

Tierärztliche Hochschule Hannover

**Abschätzen der diagnostischen Testgenauigkeit ohne
Goldstandard bei bedingt abhängigen Tests**

**Methodische Entwicklung von Modellen der latenten
Klassenanalyse**

INAUGURAL – DISSERTATION
zur Erlangung des Grades einer Doktorin
der Naturwissenschaften
- Doctor rerum naturalium -
(Dr. rer.nat.)

vorgelegt von

Clara Schoneberg

Achim

Hannover 2022

Wissenschaftliche Betreuungsgruppe: Prof. Dr. Lothar Kreienbrock

Institut für Biometrie, Epidemiologie und
Informationsverarbeitung

Tierärztliche Hochschule Hannover

Prof. Dr. Martin Ganter

Klinik für kleine Klauentiere und forensische
Medizin und Ambulatorische Klinik

Tierärztliche Hochschule Hannover

Prof. Dr. Katja Ickstadt

Fakultät Statistik

Technische Universität Dortmund

1. Gutachten Prof. Dr. Lothar Kreienbrock

Prof. Dr. Martin Ganter

Prof. Dr. Katja Ickstadt

2. Gutachten Prof. Dr. Hans-Peter Piepho

Institut für Kulturpflanzenwissenschaften

Universität Hohenheim

Tag der mündlichen Prüfung: 03.06.2022

Meiner Familie

Dieser Dissertation liegen zwei Publikationen zugrunde:

Schoneberg C, Kreienbrock L, Campe A

An Iterative, Frequentist Approach for Latent Class Analysis to Evaluate Conditionally Dependent Diagnostic Tests.

Front. Vet. Sci. 2021, 8:588176.

Schoneberg C, Böttcher J, Janowetz B, Rostalski A, Kreienbrock L, Campe A

An intercomparison study of ELISAs for the detection of porcine reproductive and respiratory syndrome virus – evaluating six conditionally dependent tests.

PLoS ONE 2022, 17(1)

**Ergebnisse dieser Dissertation wurden auf folgenden
Fachkonferenzen präsentiert:**

Schoneberg C, Kreienbrock L, Campe A

Ein iterativer, frequentistischer Ansatz der latenten Klassenanalyse für die Evaluation bedingt abhängiger diagnostischer Tests. (Poster)

DVG DACH-Epidemiologietagung 2021, Bern (Schweiz), 1.-3. September 2021.

Schoneberg C, Böttcher J, Janowetz B, Kreienbrock L, Campe A

Evaluation der diagnostischen Güte von sechs ELISA Tests zum Nachweis von PRRS mit einem Modell der latenten Klassenanalyse. (Vortrag)

Jahrestagung der DVG-Fachgruppe „AVID“ 2021, online, 14. September 2021

Inhaltsverzeichnis

Inhaltsverzeichnis	vii
Abkürzungsverzeichnis	ix
1 Einleitung	1
1.1 Notwendigkeit diagnostischer Testevalution	1
1.2 Biologische Problemstellung.....	2
1.3 Statistische Lösungsansätze	3
1.4 Zielstellung.....	5
1.5 Aufbau der Arbeit.....	5
2 Publikationen.....	7
2.1 Publikation I: An iterative, frequentist approach for latent class analysis to evaluate conditionally de-pendent diagnostic tests.....	7
2.2 Publikation II: An intercomparison study of ELISAs for the detection of porcine reproductive and respiratory syndrome virus – Evaluating six conditionally dependent tests.....	19
3 Übergreifende Diskussion.....	37
3.1 Herausforderungen in der Anwendung der Methode	37
3.1.1 Stichprobengröße	37
3.1.2 Startwerte	39
3.1.3 Anzahl der verwendeten Tests	40
3.1.4 Stabilität der Schätzung.....	42
3.1.5 Parameterschranken	43
3.2 Anwendungsbereich der Methode.....	47
3.2.1 Anwendungsbereich in der medizinischen diagnostischen Testevalution	47
3.2.2 Biologische Restriktionen	48
3.2.3 Weitere Anwendungsbereiche.....	49
3.2.4 Technische Anwendbarkeit (Softwarerealisation)	50
3.3 Vergleich mit Verfahren der Bayes-Statistik	51
3.4 Weiterentwicklung der Methode	54
3.5 Übergreifende Schlussfolgerungen und Ausblick.....	56
4 Zusammenfassung.....	59

5	Summary.....	61
6	Literaturverzeichnis.....	63
	Anhänge.....	67
	Danksagung.....	103

Abkürzungsverzeichnis

<i>C. burnetii</i>	<i>Coxiella burnetii</i>
DNA	Desoxyribonukleinsäure (<i>deoxyribonucleic acid</i>)
ELISA	<i>Enzyme-linked Immunosorbent Assay</i>
EM	Erwartungs-Maximierung (<i>expectation-maximization</i>)
LCA	Latente Klassenanalyse (<i>latent class analysis</i>)
PCR	Polymerase Kettenreaktion (<i>polymerase chain reaction</i>)
PRRS	Porzines Reproduktives und Respiratorisches Syndrom

1 Einleitung

1.1 Notwendigkeit diagnostischer Testevaluation

Infektionen finden durch die Aufnahme von Krankheitserregern, wie zum Beispiel Viren oder Bakterien, statt und stellen eine lokale oder allgemeine Störung des Organismus eines Individuums dar. Diese Störungen des Organismus können bei dem Individuum klinisch unauffällig sein, so dass eine Erregerausscheidung durch das Individuum übersehen wird, oder mit unspezifischen Symptomen einhergehen, die einem bestimmten Erreger nicht eindeutig zugeordnet werden können (Li 2018). Um bestimmte Infektionen nachweisen zu können und somit die richtige Behandlung anzuwenden und eine Erregerübertragung auf andere Individuen zu verhindern, wird also ein diagnostischer Test benötigt, der ein klinisches oder labormedizinisches Verfahren zum Nachweis der Erkrankung darstellt (Greiner 2003).

Jeder diagnostische Test hat grundsätzlich jedoch einen Anteil falscher Ergebnisse. Verschiedene Einflussfaktoren wie zum Beispiel eine falsche Anwendung, Kreuzreaktionen oder eine zu geringe Erregermenge in der Probe können zu falsch-positiven oder falsch-negativen Ergebnissen führen. Dieser Anteil ist abhängig vom verwendeten Testverfahren und Probenmaterial sowie von der untersuchten Population und kann daher stark variieren (Gardner et al. 2019).

Abhängig vom Anteil falscher Ergebnisse und somit der Sensitivität und Spezifität gibt es unterschiedliche Anwendungsgebiete für einen diagnostischen Test, denn dieser muss passend zum Ziel der Anwendung nach dem „fit-for-purpose Konzept“ gewählt werden (World Organisation for Animal Health 2018). Ein Screening Test zum Beispiel sollte eine sehr hohe Sensitivität besitzen, um möglichst viele positive Ergebnisse richtig zu erkennen. Ein Bestätigungs- test hingegen sollte eine hohe Spezifität haben, um falsch-positive Ergebnisse herauszufiltern und so unnötige Interventionen zu verhindern.

Auch in epidemiologischen Untersuchungen, in denen der Anteil positiver Individuen in einer Population bestimmt werden soll, muss die Testgüte des verwendeten Tests berücksichtigt werden. Fehlerhafte Testergebnisse führen zu einem verzerrten Prävalenzwert der untersuchten Population. Die gemessene Prävalenz kann korrigiert werden, indem anhand der angegebenen Werte für die Sensitivität und Spezifität des Tests der Anteil falscher Diagnosen bestimmt und dieser der Prävalenz hinzufügt beziehungsweise abgezogen wird (Rogan und Gladen 1978). So kann die beobachtete (apparente) Prävalenz einer Population unabhängig vom gewählten Test adjustiert und damit der wahren Prävalenz angenähert werden, wenn

dessen Testgüte bekannt ist. Ein diagnostischer Test muss also zunächst evaluiert werden, um seine Werte für die Sensitivität und Spezifität zu bestimmen. Mit diesen Informationen kann das Anwendungsgebiet festgelegt und die Prävalenz korrigiert werden.

1.2 Biologische Problemstellung

Eine häufig angewendete Evaluationsmethode für einen diagnostischen Test ist der Vergleich mit einem Goldstandard, also einem Test, welcher die gleiche Erkrankung mit optimaler Testgüte nachweist. Anhand der voneinander abweichenden Resultate wird der Anteil falscher Ergebnisse des untersuchten Tests bestimmt und daraus die Sensitivität sowie die Spezifität berechnet. Dieser Ansatz setzt einen optimalen Goldstandardtest voraus, den es nicht immer gibt. Wird ein Test anhand eines suboptimalen Goldstandards beurteilt, sind die berechneten Ergebnisse verzerrt. Eine weitere Möglichkeit der diagnostischen Testevalution ist die Anwendung des Tests auf eine Goldstandardpopulation mit bekanntem Krankheitsstatus. So kann ein falsches Ergebnis direkt festgestellt und die Testgüte berechnet werden. Jedoch kann die gewählte Population aufgrund ihrer spezifischen Eigenschaften die Werte der Sensitivität und Spezifität beeinflussen. Die Werte der Testgüten wären somit nicht auf andere Populationen anwendbar (Dohoo et al. 2009).

Die latente Klassenanalyse (LCA) ist eine Evaluationsmethode, welche auf jede Population auch ohne einen Goldstandardtest anwendbar ist (Walter und Irwig 1988). Sie basiert auf der Idee, dass diagnostische Tests einen nicht direkt feststellbaren, also latenten Krankheitsstatus messen. Jeder dieser Tests hat einen individuellen Anteil falsch-positiver und falsch-negativer Ergebnisse. Werden mehrere diagnostische Tests auf dieselbe Population angewendet, kann der Krankheitsstatus jedes Individuums anhand des so entstehenden Antwortmusters bestimmt werden. Alle Individuen werden somit einer latente Klasse für den positiven beziehungsweise negativen Krankheitsstatus zugeordnet. Durch den Anteil der abweichenden Ergebnisse jedes Tests innerhalb der beiden latenten Klassen wird dann die Testgüte berechnet.

Eine wichtige Voraussetzung der LCA ist die bedingte Unabhängigkeit der Tests. Diese setzt voraus, dass für alle Individuen mit demselben latenten Krankheitsstatus die Ergebnisse eines Tests keine Rückschlüsse auf das Ergebnis eines anderen Tests zulassen. Werden beispielsweise eine Polymerase-Kettenreaktion (PCR) zum direkten Nachweis der Desoxyribonukleinsäure (DNA) eines Bakteriums und ein *Enzyme-linked Immunosorbent*

Assay (ELISA) für den Nachweis von Antikörpern verwendet, prüfen beide Methoden unterschiedliche Formen der Infektion. Ist die Erregermenge sehr gering oder werden eventuell nicht die optimalen Sequenzen zum Nachweis des Erregers verwendet, beeinflusst dies das Ergebnis der PCR. Auf den Nachweis von Antikörpern hat dies jedoch keinen Einfluss. Andersherum hat eine falsche Reaktion auf den Antikörper keinen Einfluss auf den direkten Erreger nachweis (Parija 2013). Somit gibt ein falsches Ergebnis des einen Tests keinen Hinweis auf das Ergebnis des anderen Tests.

Werden hingegen zum Nachweis derselben Antikörper zwei verschiedene ELISA-Tests angewendet, nutzen beide Tests dasselbe biologische Prinzip. Kommt es dann bei einem ELISA aufgrund von Kreuzreaktionen beziehungsweise unspezifischen Reaktionen zu einem falsch-positiven Resultat, können dieselben Reaktionen beim anderen ELISA auftreten (Virella 2001). Auch eine zu geringe Antikörperkonzentration in der Probe kann beide Tests beeinflussen. Somit erhöht das falsche Ergebnis des einen Tests die Wahrscheinlichkeit einer falschen Diagnose des anderen Tests und die Tests sind nicht mehr bedingt unabhängig voneinander.

Werden diese Tests nun gemeinsam in einer LCA anwendet, stimmen die bedingt abhängigen Tests stärker miteinander überein als es ihre Testgüten rechtfertigen würden. Somit berechnet das latente Klassenmodell anhand der höheren Übereinstimmung eine zu hohe Testgüte dieser beiden Tests und eine zu geringe Testgüte eventueller zusätzlich betrachteter bedingt unabhängiger Tests. Die Ergebnisse wären verzerrt und die LCA nicht anwendbar (Vacek 1985).

1.3 Statistische Lösungsansätze

Es gibt bereits einige Methoden, die keinen Goldstandardtest beziehungsweise keine Goldstandardpopulation benötigen und in der Lage sind, bedingte Abhängigkeiten diagnostischer Tests zu berücksichtigen. Diese Methoden basieren auf zwei unterschiedlichen Zweigen der Statistik: der Bayesschen Statistik und der klassischen, frequentistischen Statistik.

Die Bayesschen Methoden betrachten die Parameter, also die Prävalenz, die Testgüten sowie die bedingten Abhängigkeiten, als Zufallsvariablen. Dabei wird eine A-priori-Verteilung für jeden Parameter festgelegt, die das Vorwissen zur Verteilung des Parameters darstellt. Diese Verteilungen und die beobachteten Daten werden verwendet, um anhand der Formel von

Bayes die A-Posteriori-Verteilungen für die Parameter zu berechnen, aus denen wiederum Punktschätzer abgelesen werden können. Bayessche Modelle benötigen daher zwingend eine A-priori-Verteilung (Held 2008).

Die frequentistischen Lösungsmethoden hingegen berechnen die Parameterwerte ausschließlich basierend auf den vorhandenen Daten, zum Beispiel mittels einer Maximum-Likelihood-Schätzung. Es werden die Parameterwerte ausgewählt, welche die beobachteten Daten mit maximaler Wahrscheinlichkeit unter dem gegebenen Modell erzeugt haben (Held 2008). Dabei werden die Prävalenz, die Testgüten und die bedingten Abhängigkeiten als feste Werte angesehen, die anhand der beobachteten Daten abgeschätzt werden können.

Für den Bayesschen Ansatz der bedingt abhängigen latenten Klasseanalyse wurden bereits verschiedene Modelle entwickelt. Ein Modell betrachtet die Kovarianzen zwischen den Tests als fixe Effekte; ein zweites Modell berücksichtigt die Variation der Testparameter als Zufallseffekte (Dendukuri und Joseph 2001). In einem anderen Ansatz wird eine einfache Methode für die Berücksichtigung einer Korrelation zweier diagnostischer Tests vorgestellt, die mindestens zwei Populationen mit unterschiedlichen Prävalenzen benötigt (Georgiadis et al. 2003). Darüber hinaus wurden drei verschiedene Modelle für unterschiedliche Formen der bedingten Abhängigkeit beschrieben (Black und Craig 2002).

Als frequentistische Ansätze der bedingt abhängigen latenten Klasseanalyse wurden ein latentes Randmodell (Yang und Becker 1997), verschiedene Zufallseffektmodelle (Qu et al. 1996; Albert et al. 2001; Zhang et al. 2012) und ein verallgemeinertes Mischungsmodell (Liu et al. 2015) entwickelt. Ein weiterer Ansatz verwendet log-lineare und logistische Modelle, um bedingte Abhängigkeiten zu untersuchen (Hanson et al. 2000).

Die obigen Lösungen frequentistischer Ansätze sind grundsätzlich in der Lage, bedingte Abhängigkeiten zu berücksichtigen, weisen jedoch einige Schwächen auf, welche die praktische Anwendbarkeit auf die diagnostische Testevaluation erschweren. So benötigen zum Beispiel viele dieser Modelle mindestens vier verschiedene diagnostische Tests oder mindestens zwei unterschiedliche Populationen oder einen zumindest teilweise bekannten Krankheitsstatus der untersuchten Individuen, um reguläre Lösungen zu bestimmen.

Bayesschen Ansätze kommen dagegen zum Beispiel nur mit einer Population aus, setzen zur Berechnung der Parameterwerte aber Vorwissen in Form von A-priori-Verteilungen, die den Grad der Sicherheit dieser Vorinformationen widerspiegeln, voraus.

Somit stellen sowohl frequentistische als auch Bayessche Verfahren Anforderungen, welche zu Schwierigkeiten bei der praktischen Anwendbarkeit führen können.

1.4 Zielstellung

Das Ziel dieser Arbeit ist es, eine Methode zur Evaluation bedingt abhängiger diagnostischer Tests ohne Goldstandard zu entwickeln. Sie sollte ohne umfassendes Vorwissen und ohne Formulierung einer A-priori-Verteilung anwendbar sein. Zusätzlich besteht das Ziel, eine Methode zu entwickeln, die für die Analyse von Datensätzen mit mindestens drei Tests angewendet auf eine einzelne Population geeignet ist. Die Präzision der Schätzung sollte dann anhand von simulierten und realen Daten bestimmt werden.

1.5 Aufbau der Arbeit

Zunächst wird in Publikation I ein latentes Klassenmodell für bedingt abhängige Tests hergeleitet und der zugehörige frequentistische Algorithmus vorgestellt. Anhand verschiedener Simulationsszenarien aus dem Bereich der Diagnostik wird der Algorithmus evaluiert und die Genauigkeit seiner Schätzungen der klassischen LCA und der Bayesschen Methode für bedingt abhängige Tests gegenübergestellt. Der neu entwickelte Ansatz benötigt gut gewählte Startwerte für die Parameter, um zu den richtigen Werten zu konvergieren. Unter der Annahme gleichen Vorwissens führt die neue Methode in den betrachteten Simulations-szenarien zu besseren Ergebnissen als die klassische LCA. Demgegenüber schätzt die Bayessche Methode die Parameter ähnlich gut ab. Dadurch wird die grundsätzliche Anwendbarkeit der Methode bestätigt. Die vollständigen Daten aus Publikation I sind in Anhang 1 abgebildet.

In Publikation II wird der iterative, frequentistische Ansatz der latenten Klassenanalyse auf einen realen Datensatz angewendet. Anhand der Daten von 812 untersuchten Schweinen werden sechs verschiedene ELISAs evaluiert, die auf das Porzine Reproduktive und Respiratorische Syndrom (PRRS) testen. Es werden verschiedene Informationsquellen genutzt, um geeignete Startwerte für den Algorithmus zu finden. Diese werden jeweils auf den gesamten Datensatz sowie eine geimpfte Untergruppe angewendet. Es wird gezeigt, dass der neu entwickelte Ansatz der latenten Klassenanalyse für den gesamten Datensatz zu guten und übereinstimmenden Ergebnissen führt. Bei der geimpften Untergruppe gibt es jedoch

stark abweichende Ergebnisse, was auf eine erschwerte Schätzung bei sehr homogenen Gruppen schließen lässt. Anhand dieses Beispiels wird daher gezeigt, dass die vorgestellte Methode auf reale Datensätze mit bedingt abhängigen Tests anwendbar ist, jedoch noch genauere Kenntnisse zur untersuchten Population benötigt werden. Die vollständigen Daten aus Publikation II sind in Anhang 2 abgebildet.

In der übergreifenden Diskussion werden zunächst im Kapitel 3.1 technische Herausforderungen für die praktische Anwendung der Methode betrachtet. Die benötigte Stichprobengröße wird abgeschätzt und die Wahl geeigneter Startwerte für den entwickelten Algorithmus wird diskutiert. Anschließend werden Faktoren analysiert, welche die Stabilität des Verfahrens sowie die Güte der Schätzung beeinflussen und folglich bei der Anwendung auf reale Datensätze zu berücksichtigen sind. Danach wird die für das Modell notwendige Anzahl der diagnostischen Tests untersucht. Abschließend wird auf Schwierigkeiten der in Publikation I vorgeschlagenen Parametergrenzen zur Verbesserung der Schätzung der Methode eingegangen, eine alternative Lösung vorgestellt und deren Vor- und Nachteile diskutiert. Kapitel 3.2 beschreibt den Anwendungsbereich der vorgestellten Methode. Dabei werden die Anforderungen an die verwendeten Tests und die Population dargestellt und weitere Anwendungsbiete der neuen Methode auch abseits der (veterinär-)medizinischen Diagnostik betrachtet. Im Anschluss wird die praktische Anwendbarkeit des Algorithmus bewertet. In Kapitel 3.3 wird das in dieser Arbeit entwickelte Modell mit dem Bayesschen latenten Klassenmodell verglichen und ihre jeweiligen möglichen Anwendungsbereiche und Einschränkungen gegenübergestellt. Danach wird in Kapitel 3.4 illustriert inwiefern es Potenzial gibt, den Algorithmus anzupassen, so dass die Schätzung robuster wird und mehr Anwendungsfälle berücksichtigt werden können. Schließlich wird in Kapitel 3.5 herausgestellt, welche Erkenntnisse aus dieser Arbeit neben den Ergebnissen der oben genannten Veröffentlichungen entstanden sind und wie die neu entwickelte Methode in Zukunft in der Forschung angewendet werden kann.

2 Publikationen

2.1 Publikation I: An iterative, frequentist approach for latent class analysis to evaluate conditionally dependent diagnostic tests

Schoneberg C, Kreienbrock L and Campe A (2021) An Iterative, Frequentist Approach for Latent Class Analysis to Evaluate Conditionally Dependent Diagnostic Tests. *Front. Vet. Sci.* 8:588176. doi: 10.3389/fvets.2021.588176

Journal Frontiers in Veterinary Science

Editor Beate Pinior (Veterinärmedizinische Universität Wien, Österreich)

Eingereicht 28. Juli 2020

Akzeptiert 07. Januar 2021

Veröffentlicht 10. Februar 2021



An Iterative, Frequentist Approach for Latent Class Analysis to Evaluate Conditionally Dependent Diagnostic Tests

Clara Schoneberg*, Lothar Kreienbrock and Amely Campe

Department of Biometry, Epidemiology and Information Processing, WHO Collaborating Centre for Research and Training for Health in the Human-Animal-Environment Interface, University for Veterinary Medicine Hannover, Hannover, Germany

OPEN ACCESS

Edited by:
Beate Pinior,

University of Veterinary Medicine
Vienna, Austria

Reviewed by:
Brecht Devleesschauwer,
Sciensano, Belgium
Chiara Antonini,
ICT4life, Italy

***Correspondence:**
Clara Schoneberg
clara.schoneberg@taho-hannover.de

Specialty section:

This article was submitted to
Veterinary Epidemiology and
Economics,
a section of the journal
Frontiers in Veterinary Science

Received: 28 July 2020

Accepted: 07 January 2021
Published: 10 February 2021

Citation:

Schoneberg C, Kreienbrock L and
Campe A (2021) An Iterative,
Frequentist Approach for Latent Class
Analysis to Evaluate Conditionally
Dependent Diagnostic Tests.
Front. Vet. Sci. 8:588176.
doi: 10.3389/fvets.2021.588176

Latent class analysis is a well-established method in human and veterinary medicine for evaluating the accuracy of diagnostic tests without a gold standard. An important assumption of this procedure is the conditional independence of the tests. If tests with the same biological principle are used, this assumption is no longer met. Therefore, the model has to be adapted so that the dependencies between the tests can be considered. Our approach extends the traditional latent class model with a term for the conditional dependency of the tests. This extension increases the number of parameters to be estimated and leads to negative degrees of freedom of the model, meaning that not enough information is contained in the existing data to obtain a unique estimate. As a result, there is no clear solution. Hence, an iterative algorithm was developed to keep the number of parameters to be estimated small. Given adequate starting values, our approach first estimates the conditional dependencies and then regards the resulting values as fixed to recalculate the test accuracies and the prevalence with the same method used for independent tests. Subsequently, the new values of the test accuracy and prevalence are used to recalculate the terms for the conditional dependencies. These two steps are repeated until the model converges. We simulated five application scenarios based on diagnostic tests used in veterinary medicine. The results suggest that our method and the Bayesian approach produce similar precise results. However, while the presented approach is able to calculate more accurate results than the Bayesian approach if the test accuracies are initially misjudged, the estimates of the Bayesian method are more precise when incorrect dependencies are assumed. This finding shows that our approach is a useful addition to the existing Bayesian methods, while it has the advantage of allowing simpler and more objective estimations.

Keywords: conditional dependence, sensitivity, specificity, maximum likelihood, veterinary medicine

INTRODUCTION

Information about the occurrence of livestock diseases in an animal population is important in many applications, such as surveillance and vaccination programs or the verification of the freedom from the disease. Therefore, the disease status of the individual animals is assessed using a diagnostic test. However, every test also has a number of incorrect results, which, depending on

the disease, may have serious economic, social or political consequences. This result can be avoided by sequentially examining a subset of the animals with a different test or by testing all the animals with multiple diagnostic tests from the outset and thus confirming the diagnosis (1).

The latter was the chosen approach in a field trial to determine the prevalence of *Brucella abortus* in cattle in Northern Ireland in 2003/2004 (2). Each animal was examined with six serological tests simultaneously, which yielded conflicting results for a large number of the animals. For example, there were two positive and four negative test results for one animal. Knowledge of the test accuracy and therefore the probability of a false diagnosis of the tests are the basis for assessing the true underlying disease status. This information is obtained by evaluating the diagnostic tests used. To assess the sensitivity and specificity, the diagnostic test is usually compared with a gold standard test or is applied to animals with known disease status (1).

However, the test accuracy depends on many biological factors, such as the animal species, race, sex and immune history. For this reason, the diagnostic test accuracy varies across populations, and the values obtained in clinical evaluation studies are only conditionally applicable to the field settings (3). Therefore, if the exact values of the test accuracy for a set of diagnostic tests in a field study are of interest and a gold standard or animals with known disease status are not available, a latent class model can be used (4).

In this context, latent class analysis (LCA) is based on the assumption that observed categorical indicators imperfectly measure an underlying latent structure. By sampling the values of the categorical variable for a set of observations, the method is able to discover the latent structure and the error in the indicators. Applying this principle to the field of diagnostic test evaluation, the true unknown disease status is measured by observed diagnostic tests. By analyzing the response pattern of a set of tests, the prevalence of the disease in the sample and the diagnostic accuracy of every test used in this model can be discovered. Hence, the specific test performance under the given conditions such as the study settings and the structure of the subpopulation can be estimated.

The Bayesian approach to latent class models of the test accuracy is widely used in veterinary medicine (5). However, latent class analyses can also be implemented in a frequentist framework, and the model parameters can be estimated by classical inference (6, 7). This method does not require any prior distribution and therefore allows an easy and objective estimation of the parameters. A prior distribution can be a benefit for the estimation if there is much information that can be incorporated into the model. However, when new or less established tests are used, there is limited previous knowledge and it is difficult to determine an informative prior distribution. The Bayesian model requires also a burn-in phase and is often more time consuming, since the convergence and therefore the termination condition are difficult to assess. Consequently, it seems reasonable to use classic frequentist methods, especially when the user has little or no prior knowledge of the distributions due to the biological methods used. Regardless of the chosen method, the basis for most latent class models is that (i) the tested individuals can be divided into two or more populations

with different prevalence values; (ii) the tests have the same sensitivity and specificity in all populations; and (iii) the tests are conditionally independent (i.e., given the true disease status). These assumptions are known as the Hui-Walter paradigm (6). However, in many cases, the assumption of conditional independence is hard to justify, especially if the tests are based on the same biological principle, such as the detection of antibodies. Ignoring the dependency structure among the diagnostic tests will introduce bias in the estimates so that a positive association will overestimate the test properties while a negative association will lead to an underestimation of the test accuracy (8).

Some Bayesian methods allow the consideration of conditionally dependent tests. A fixed effects model and a random effects model for data from a single population were developed in one approach (9), while another study used a simple method for two diagnostic tests that can be applied when two or more populations are studied (10). Moreover, three different models were described for varying forms of conditional dependence (11).

There are also some frequentist approaches for incorporating a dependence structure into the latent class analysis. A latent marginal model (12), different random effect models (13–15) and a more generalized mixture model (16) have been proposed. Another approach uses log-linear and logistic models to investigate conditional dependence (17). However, this model assumes that the true disease status of each individual is available, which is not possible in most diagnosis applications.

Although the solutions addressed above are available, we propose a frequentist method for estimating the prevalence, diagnostic test accuracy and dependence structure because we would like to present an easy-to-apply approach even for situations with no accessible prior information. The solution was intended to fit even when only three diagnostic tests are available and the status of each individual is unknown. We present the model as well as the algorithm and discuss its performance in different simulated scenarios, which were adopted from real-world examples in veterinary medicine to examine the performance of our method under different circumstances. The non-mathematically inclined reader may skip the following three subsections describing the statistical model.

MATERIALS AND METHODS

The Latent Class Model

In a latent class model, it is assumed that there is a latent variable with C classes. The proportion of each class is estimated by M observed variables. Let the vector $Y_i = (Y_{i1}, \dots, Y_{iM})$, $i = 1, \dots, N$ represent individual i 's response pattern with the possible values $0, \dots, r_m$ for observation Y_{im} . The probability of membership in the latent class c can be expressed as γ_c with $\sum_{c=0}^{C-1} \gamma_c = 1$, and the probability of the response r_m to variable m in class c can be expressed as $\rho_{m,r_m|c}$. Let $I(\cdot)$ be the indicator function. Then, the likelihood of parameters γ and ρ for the observations y is denoted by Formula (a).

$$L(\gamma, \rho | Y) = \prod_{i=1}^N \prod_{c=0}^{C-1} \gamma_c^{\sum_{m=1}^{R_m} I(Y_{im}=r_m)} \rho_{m,r_m|c}^{I(Y_{im}=r_m)} \quad (a)$$

In this publication, we discuss the application to the diagnostic test evaluation for dichotomous diagnoses (i.e., with/affected vs. without/not affected). Thus, we only allow dichotomous responses for the observed variables and for two latent classes. This results in the simplified likelihood, which can be written as Formula (b).

$$L(\gamma, \rho | Y) = \sum_{i=1}^N \sum_{c=0}^1 \gamma_c \prod_{m=1}^M \prod_{r_m=0}^1 \rho_{m, r_m | c}^{I(y_{im} = r_m)} \quad (\text{b})$$

It differs from the two-test, two-population scenario originally proposed by Hui and Walter (6) only in the number of populations and tests used. However, due to this parameter changes there are no closed-form maximum-likelihood solutions for this model and the values are estimated iteratively.

Key assumption of the approach is the conditional independence (i.e., given the true disease status) of all the tests (6). Therefore, within a latent class, the result of one test does not give any indication of the result of the other tests. If there are diagnostic tests with the same biological principle used in the analysis, that assumption no longer holds because the same external factors influence their diagnoses. As ignoring this underlying dependence structure leads to biased results (8), a term to describe the dependency must be included when using tests with the same (biological) testing principle.

Conditional Dependence—The Interaction Term

If diagnostic tests are independent, the conditional response probabilities result from the product of the tests' individual response probabilities as described in the likelihood (b). It can be written in the simplified term (c) for three diagnostic tests i , j , and k .

$$\rho_{i, r_i | c} \cdot \rho_{j, r_j | c} \cdot \rho_{k, r_k | c} = P(r_i, r_j, r_k | c) \quad (\text{c})$$

However, if the tests are conditionally dependent on the underlying disease status, these dependencies must be taken into account when calculating the conditional item response probabilities. They are determined within the two latent classes containing the observations with a positive and a negative disease status, respectively.

The pairwise dependencies η_{ij}^c of tests i and j conditional on the latent class c are estimated by comparing the observed matching correct response patterns $P(c, c | c)$ with the test accuracies of the diagnostic tests used. If the observed test agreement is stronger than expected based on the known test accuracy, there is a positive dependency between the tests that cannot be attributed to the underlying disease status alone. If the test agreement is weaker than expected, there is a negative dependency. Formula (d) denotes the dependency of the diagnostic tests conditional on the latent class of the not affected/diseased animals $c = 0$, with Sp_i and Sp_j denoting the known specificities of tests i and j .

$$\eta_{ij}^0 = P(0, 0 | 0) - Sp_i Sp_j \quad (\text{d})$$

The dependency of three tests i, j and k can be estimated analogously by calculating the difference between their observed test agreement $P(c, c, c | c)$ and the expected proportion of correct observations in class c . However, since the tests are pairwise dependent on each other, these values must be taken into account, as they are expected to influence the result. Due to their dependency, matching test results have an increased item response probability (i.e., prevalence) compared to the independent cases whereas the probabilities of the deviating test results decrease. This means that the pairwise dependencies must be added to the product of the test accuracies to calculate the expected test agreement. In addition, a possible dependency of the pairwise dependent tests on the third test is considered by multiplying the dependency term by the test accuracy of the third test. This results in the conditional dependency of three tests in the latent class of the not affected/diseased animals $c = 0$ to be calculated as described in Formula (e).

$$\eta_{ijk}^0 = P(0, 0, 0 | 0) - (Sp_i Sp_j Sp_k + \eta_{ij}^0 Sp_k + \eta_{ik}^0 Sp_j + \eta_{jk}^0 Sp_i) \quad (\text{e})$$

The conditional dependencies for the class of the in-/affected animals $c = 1$ can be calculated analogously.

Some authors use the terms “dependency” and “correlation” interchangeably (18, 19). We prefer “dependency” and therefore use it in this publication because the related term calculates concordance rather than a traditional correlation of quantitative variables.

The latent disease status determines the correct diagnosis of an observation, while other external factors trigger a misdiagnosis. Thus, only matching incorrect results are of interest to assess the dependency of the tests. The proportion of incorrect results, i.e., the proportion of incorrectly diagnosed animals, is determined by the accuracy of the test, which causes specific restrictions for the dependency parameter settings (20). This means, the higher the accuracies, the greater the agreement due to the correct diagnosis and the smaller the possible dependency. For example, if both tests have test accuracies of 100%, then all of their results match and their conditional dependency is zero. Therefore, the test accuracy of the examined tests determines the possible range of the dependencies.

For the dependencies to be comparable, they have to be detached from the test accuracies. In case of pairwise dependencies, this is achieved by standardizing Formula (d) to Formula (f). The three-test dependencies are standardized analogously.

$$Z_{ij}^0 = \frac{\eta_{ij}^0}{\sqrt{Sp_i(1 - Sp_i)} \sqrt{Sp_j(1 - Sp_j)}} \quad (\text{f})$$

This model is applicable in situations when the results of at least two diagnostic tests are available (21). However, here, we focus on the three-test model.

The dependencies of the diagnostic tests are calculated within both latent classes and remain constant for all possible combinations of the results of the three tests in the respective class. Therefore, these values can also be determined by using all observed response patterns. Only the signs of the dependencies

have to be adjusted due to matching or differing test results. Changing these equations results in the functions (g) of the conditional items response probabilities for the class of non-infected animals $c = 0$.

$$\begin{aligned}
 P(0, 0, 0|0) &= Sp_1 Sp_2 Sp_3 + \eta_{12}^0 Sp_3 + \eta_{13}^0 Sp_2 + \eta_{23}^0 Sp_1 + \eta_{123}^0 \\
 P(0, 0, 1|0) &= Sp_1 Sp_2 (1 - Sp_3) + \eta_{12}^0 (1 - Sp_3) - \eta_{13}^0 Sp_2 \\
 &\quad - \eta_{23}^0 Sp_1 - \eta_{123}^0 \\
 P(0, 1, 0|0) &= Sp_1 (1 - Sp_2) Sp_3 - \eta_{12}^0 Sp_3 + \eta_{13}^0 (1 - Sp_2) \\
 &\quad - \eta_{23}^0 Sp_1 - \eta_{123}^0 \\
 P(0, 1, 1|0) &= Sp_1 (1 - Sp_2) (1 - Sp_3) - \eta_{12}^0 (1 - Sp_3) \\
 &\quad - \eta_{13}^0 (1 - Sp_2) + \eta_{23}^0 Sp_1 + \eta_{123}^0 \\
 P(1, 0, 0|0) &= (1 - Sp_1) Sp_2 Sp_3 - \eta_{12}^0 Sp_3 - \eta_{13}^0 Sp_2 \\
 &\quad + \eta_{23}^0 (1 - Sp_1) - \eta_{123}^0 \\
 P(1, 0, 1|0) &= (1 - Sp_1) Sp_2 (1 - Sp_3) - \eta_{12}^0 (1 - Sp_3) + \eta_{13}^0 Sp_2 \\
 &\quad - \eta_{23}^0 (1 - Sp_1) + \eta_{123}^0 \\
 P(1, 1, 0|0) &= (1 - Sp_1) (1 - Sp_2) Sp_3 + \eta_{12}^0 Sp_3 - \eta_{13}^0 (1 - Sp_2) \\
 &\quad - \eta_{23}^0 (1 - Sp_1) + \eta_{123}^0 \\
 P(1, 1, 1|0) &= (1 - Sp_1) (1 - Sp_2) (1 - Sp_3) + \eta_{12}^0 (1 - Sp_3) \\
 &\quad + \eta_{13}^0 (1 - Sp_2) - \eta_{23}^0 (1 - Sp_1) - \eta_{123}^0
 \end{aligned} \tag{g}$$

The item-response probabilities conditional on the latent class of the affected/diseased animals $c = 1$ are determined analogously.

The (standardized) dependency (e) indicates the strength of the interdependence of the tests, i.e., the share of their concordant false diagnoses. To interpret this measure, the size and direction should be considered. If two tests completely agree on their incorrect diagnosis (i.e., both tests assign the incorrect disease status to exactly the same animals), then they have a (standardized) dependency of 1. If the two tests agree only at random regarding the incorrect diagnosis of the disease status, the observed agreement between these tests matches the expected agreement between two independent tests. Hence, the (standardized) dependency is zero. The dependency is negative when there are fewer matching results than expected by chance. Thus, this measure has an interpretation similar to Cohen's Kappa. However, here, negative values play a subordinate role in the application to diagnostic tests, since similar testing principles tend to lead to increased agreement in incorrect decisions. It is very unlikely that two tests with a similar test procedure have a negative dependency, as this phenomenon would imply that higher biological similarity leads to a lower level of agreement.

Conditional Dependence—Illustrative Examples

To obtain a better idea of the magnitude of the conditional dependency, published studies that contain observations with a confirmed latent status may be discussed. Although publications with the information needed are rare and provide a rather rough indication on the size of the dependencies realized under the given study conditions, they may yield a valuable starting value for subsequent analyses. The following examples

set the framework for the magnitude of the dependency in our simulation.

First, we calculated the standardized dependency from study data on toxoplasmosis in pigs (10). In that study, the dependency between a microscopic agglutination test and an ELISA was 0.33 for the positive results and 0.49 for the negative results. In another investigation, the direct detection of *Strongyloides* infection in the stools of Cambodian refugees in Canada was compared with a serological examination (22). Based on these data, we calculated dependencies of 0.18 and 0.17 for the positive and negative observations, respectively.

The Algorithm

In the last subsections it was described how the classical frequentist latent class analysis can be extended by a term to describe the conditional dependencies between the diagnostic tests. Due to the general misspecification of the describing parameters within the setting of conditional dependency, an iterative algorithm is proposed. Here, we present a solution for the use of three diagnostic tests. The basic idea of the algorithm is to consider alternately the test accuracies and the conditional dependencies as fixed values. Thus, the method presented here has the advantage of always resulting in a positive number of degrees of freedom in each iteration step for three tests in one population compared to other methods for the estimation of test accuracies for conditionally dependent tests [e.g., (10, 12, 13, 21)]. The algorithm includes the following steps:

- (i) Choose suitable starting values for the test accuracies and the conditional dependencies between the tests.
- (ii) Consider the conditional dependencies as fixed. Execute the expectation maximization (EM) algorithm to estimate the best-fit test accuracies for the data. For this step, we followed the EM algorithm described in a conditional independent latent class approach (7) and slightly adopted it by replacing the conditional item response probabilities described in Formula (b) by the conditionally dependent item response probabilities (g) in the likelihood function.
- (iii) Recalculate the conditional dependency in two substeps:
 - (1) Use the conditional dependencies and the test accuracies from the previous step to calculate the latent class membership probabilities $P(r_1, r_2, r_3|c)$ of the observations by using Formula (g). An observation is assigned to the class for which it has the highest probability.
 - (2) With the knowledge of the latent status, determine the conditional dependency by using Formulas (d, e).
- (iv) Start again with step (ii) until the model converges, i.e., the log-likelihood of two consecutive models differs by < 0.00001 or the algorithm reaches 1,000 iterations.

We implemented the algorithm in R [version 3.5.0; (23)] by programming three main functions. The first function calculates the dependencies for fixed test accuracies and a set of observed response patterns, the second one determines the test accuracies and the prevalence for the given dependencies and response patterns (EM algorithm) and the third one combines the first two

functions by calling them alternately (see **Supplementary File 27** for complete source code).

Simulation Study—General Framework

We tested the applicability of the algorithm for three diagnostic tests in a single population by conducting a simulation study. Therefore, we took different combinations of the test accuracies, prevalence and conditional dependencies into account. These scenarios allowed an assessment of the performance of the iterative approach presented in this publication compared with that of the conditionally independent latent class analysis and the Bayesian approach for conditionally dependent tests. All the simulation scenarios are motivated by diagnostic tests used in veterinary medicine. As a small sample size leads to an increased margin of error, we simulated 10,000 observations.

We considered the following cases:

- (1) Three independent diagnostic tests with high test accuracies in a population with a moderate prevalence:

Diagnostic tests are conditionally independent if they are based on different biological principles, for example, if a tumor is detected using physical examination, medical imaging (e.g., sonography) and microscopic examination of a tissue sample. Scenarios such as this one should not cause any problems in the conditionally independent latent class analysis and result in very accurate estimates of that method. Therefore, it should provide the same results in the new approach presented in this publication as well as in models assuming conditional independent diagnostic tests. Hence, this scenario serves as basic validation for the newly fitted model.

- (2) Two highly dependent tests with low test accuracies and a third test with low dependencies and high test accuracies in a population with a high prevalence:

This scenario may be the most problematic in the conditionally independent latent class model: The two dependent tests may cause many matching results that lead to an overestimation of their test accuracies and underestimated values in the third test. This situation applies, for instance, to the diagnosis of infectious diseases with one antibody test applied to two different sample types (e.g., serum, feces, milk, etc.) and the often more accurate direct detection of the pathogen (24). By using the same test twice, the results are highly interdependent. The third test uses a different detection method and is therefore independent of the other results. This scenario may act as validation for the method proposed here to identify dependency structures in data and improve the correctness of model estimates.

- (3) Two highly dependent tests with low test accuracies and a third test with low dependencies and high test accuracies in a population with a low prevalence:

This is generally the same scenario as (2) but with a lower prevalence, which is common for many diseases. In this case, only a small proportion of the sample contains positive responses compared to the other possible response patterns. This phenomenon makes the estimation of both the prevalence and sensitivities more ambiguous and therefore more prone to errors.

- (4) Three diagnostic tests with moderate test accuracies and medium dependencies in a population with a high prevalence:

In this scenario, all three tests are conditionally dependent on each other, resulting in an overestimation of their test accuracies in the latent class model, which assumes conditional independence. For instance, three different veterinarians may perform a physical examination on the same group of animals with a suspected disease. They all have different qualifications and therefore different diagnostic sensitivities and specificities. However, experience in the same work environment with the same time and budgetary constraints can be the cause of consistent misdiagnoses (25). This scenario serves as validation for the method proposed to recognize dependency structures and calculate more accurate values.

- (5) Three diagnostic tests and a population with values for the test accuracies and the prevalence from a practical example with estimated values for the dependency structure:

To ensure that the procedure is evaluated under realistic conditions, this scenario uses the results from a prevalence study for *Brucella abortus* in cattle in Northern Ireland (2) that was mentioned in the introduction. It simulates a population at risk with a high prevalence. In the study, six different antibody tests were used from which we selected three for the simulation: an indirect ELISA, a competitive ELISA and a serum agglutination test. Since we had no information about the exact test dependencies of these tests, we based the simulations on their accuracies and biological principles. The two ELISA tests have low sensitivities, so they have a high proportion of false negative results, which may have similar causes in many cases due to similar biological principles and results in a high level of dependency in the positive latent class. As both tests have high specificities, similar detection methods have only a minor influence on the dependencies of the negative latent class. The serum agglutination test uses a different approach, but it is also based on the detection of antibodies; therefore, we assume it is only slightly dependent on the other tests.

Table 1 shows the parameter settings in the five simulated scenarios in detail. We considered only positive pairwise dependencies since negative values are not biologically justifiable.

Simulation Study—Starting Values

We applied nine different sets of starting values (6 well-chosen and 3 poorly chosen) to each of the five scenarios. As it can be assumed that prior knowledge of the applied tests, their dependency structure and the studied population is available, we considered six different sets of well-chosen “informative” starting values as follows (see **Supplementary Table 1**):

- (1) The correct values for the test accuracies, the dependency structure and the prevalence.
- (2) The correct values for the test accuracies and the prevalence; the dependency of the tests is stronger than that simulated.
- (3) The correct values for the test accuracies and the prevalence; the dependency of the tests is weaker than that simulated. The only exception is scenario 1 (independent tests): As weakening the dependency of independent tests leads to

TABLE 1 | Input parameter values for the data simulation of the five scenarios.

Scenario	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Prevalence in %	30.00	40.00	3.00	40.00	20.00
Sensitivity Test 1	90.00	90.00	90.00	80.00	72.00
Sensitivity Test 2	85.00	70.00	70.00	66.00	65.00
Sensitivity Test 3	90.00	65.00	65.00	70.00	97.00
Specificity Test 1	95.00	99.00	99.00	95.00	98.00
Specificity Test 2	95.00	80.00	80.00	85.00	99.00
Specificity Test 3	99.00	85.00	85.00	88.00	98.00
η_{12}^+ (Standardized)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.038 (0.200)	0.129 (0.600)
η_{13}^+ (Standardized)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.046 (0.250)	0.008 (0.100)
η_{23}^+ (Standardized)	0.000 (0.000)	0.121 (0.600)	0.121 (0.600)	0.087 (0.400)	0.012 (0.150)
η_{123}^+ (Standardized)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.004 (-0.050)	0.000 (0.000)
η_{12}^- (Standardized)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.016 (0.200)	0.001 (0.100)
η_{13}^- (Standardized)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.018 (0.250)	0.003 (0.150)
η_{23}^- (Standardized)	0.000 (0.000)	0.086 (0.600)	0.086 (0.600)	0.046 (0.400)	0.001 (0.100)
η_{123}^- (Standardized)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.001 (-0.050)	0.000 (0.000)

negative dependencies and negative dependencies are not biologically justifiable, another set of positive dependencies is used instead.

- (4) The correct values for the dependency structure; the test accuracies are better than those simulated, and the prevalence differs from the simulated value.
- (5) The correct values for the dependency structure; the test accuracies are poorer than those simulated, and the prevalence differs from the simulated value.
- (6) The values for the test accuracies, the dependency structure and the prevalence all differ (slightly) from the simulated data.

On the other hand, if a new diagnostic test is used, false assumptions about the underlying dependency structure, the test accuracy and even the prevalence are possible. Therefore, we also took three sets of poorly chosen starting values into consideration, that deviate greatly from the simulated values in terms of the test accuracy and the dependency structure (see **Supplementary Table 2**):

- (1) A value of 50% for all the test accuracies and the prevalence; the tests are assumed to be independent of each other.
- (2) Incorrect assumptions about which tests are dependent on each other, an incorrect ratio of the test accuracies and a prevalence that differs from the simulated value.
- (3) The results from the conditionally independent latent class analysis for the test accuracies and the prevalence as well as the incorrect dependencies.

Simulation Study—Parameter Restrictions

In some cases, there are justifiable restrictions for the resulting parameter values. As an example, negative dependencies between two diagnostic tests with the same biological testing principle are very unlikely. Another example for a justifiable restriction is to set the dependencies of known independent tests fixed to zero. Since the test accuracies are already limited to the unit interval

by the EM algorithm, further restrictions always depend on the situation and are therefore difficult to determine.

We examined the effect of parameter restrictions on the estimations of the iterative approach by repeating the calculations and adding restriction rules. As all limitations of the resulting parameter values require knowledge of the population, the disease and the diagnostic tests used in the study, they depend on the setting and are not generalizable. Thus, we focused on the most basic limitations and excluded unrealistic dependencies [standardized values <-1 or >1 ; (20)] and allowed only positive pairwise dependencies. Therefore, we modified the algorithm by adding an additional query to check whether the dependencies are within the newly defined limits. If that was not the case, they were adjusted in the next step. For this analysis, we limited ourselves to the scenario with the worst results (largest deviations from the simulated values) since these estimates needed the most improvement.

Bayesian Estimation

We also calculated the results for the five scenarios (**Table 1**) in the Bayesian framework of the presented latent class model (see **Supplementary File 28** for complete source code). For this, Formula (g) was implemented in the open source software OpenBUGS. In contrast to the presented iterative method, the algorithm requires prior information in the form of distribution functions. We used the package *rjags* in R (26) to calculate these priors. The basis for the calculation were the same starting values as in the iterative approach (**Supplementary Tables 1, 2**), but we also took a maximum possible uncertainty in the initial estimate of $\pm 10\%$ into account. OpenBUGS estimated the parameters using a Gibbs sampler, a special case of the Markov chain Monte Carlo (MCMC) algorithm (27). We ran the model with 10,000 iterations and a burn-in phase of 1,000 steps and compared the results with those of the iterative approach.

RESULTS

The five simulation scenarios consider different possible applications. Hence, we reflect on their results individually before we analyze them jointly to investigate possible differences. All results are shown in detail in **Supplementary Tables 3–26**.

Scenario 1: Three Independent Tests (Supplementary Tables 3, 8, 15, 20)

All three latent class analysis approaches were able to estimate the parameters precisely when independence was initially assumed (see **Supplementary Tables 3, 15**). However, under the assumption of dependent tests with incorrect accuracies the results deviated up to 4% from the simulated values in the iterative, frequentist approach. In the Bayesian model, these deviations increased to up to 7%. Only the first set of poorly chosen starting values resulted in incorrect estimates of the prevalence (see **Supplementary Table 8**) in the iterative, frequentist approach.

Nevertheless, overall, the results showed that all three approaches are equally applicable for evaluating independent diagnostic tests.

Scenario 2: Two Highly Dependent Tests in a Population With a High Prevalence (Supplementary Tables 4, 9, 16, 21)

The conditionally independent latent class analysis was not able to detect the connection between the tests (for none of the applied starting values) and therefore misjudged their accuracy by up to 20%. In contrast, the iterative method was able to determine the simulated parameters with only minor deviations of at most 8% for all well-chosen starting values (see **Supplementary Table 4**). However, the method yielded incorrect results for the accuracy parameters if the iterative process started with poorly chosen values (see **Supplementary Table 9**). This phenomenon was more prominent the more the ratio of the test accuracies and the dependencies differed from the simulated values. The Bayesian approach led to very similar results (see **Supplementary Tables 16, 21**). Thus, the algorithm is able to determine the correct parameter values, but initial information about the dependencies and the test accuracies is needed to choose the appropriate starting values.

Scenario 3: Two Highly Dependent Tests in a Population With a Low Prevalence (Supplementary Tables 5, 10, 17, 22)

The low prevalence in this scenario complicated the estimation. While the conditionally independent latent class analysis resulted in strongly deviating and unrealistic values for the outcome (e.g., a sensitivity of 13% for test 1), the iterative approach was mostly able to calculate the simulated values by using well-fitting starting values. Only the results for the sensitivity of test 1 posed a problem in two cases with deviations of more than 50% (see

Supplementary Table 5). The estimations of the specificities were more accurate.

This scenario resulted in the largest differences from the simulated values in the iterative approach (see **Figure 1**); therefore, we considered it for the modified algorithm with parameter restrictions.

The Bayesian approach led to better results than the frequentist method, as the deviations in the values for the sensitivity of test 1 reached a maximum of 11% for the well-chosen starting values (see **Supplementary Table 17**). In contrast, the divergences of the estimated specificity and prevalence were approximately the same size and therefore slightly worse than those from the frequentist approach. Poorly chosen starting values resulted in both methods in incorrect values (see **Supplementary Tables 10, 22**).

Scenario 4: Three Moderately Dependent Tests With a High Prevalence (Supplementary Tables 6, 11, 18, 23)

The estimates of the conditionally independent latent class analysis deviated by up to 16% from the simulated values while the iterative method and the Bayesian approach were able to obtain results that were more precise (maximum deviations of 6 or 12%; see **Supplementary Tables 6, 18**).

However, there was one exception in the Bayesian approach. Starting value set 5 (underestimated prevalence and test accuracy) led to values that deviated up to 20% from the simulated values. The initial misjudgment of the test accuracies and the prevalence therefore has a stronger effect on the Bayesian approach in this scenario than on the frequentist method.

The results of the poorly chosen starting values differed considerably from the simulated values in both methods (see **Supplementary Tables 11, 23**).

Scenario 5: Brucellosis Example (Supplementary Tables 7, 12, 19, 24)

While the conditionally independent latent class analysis overestimated the values by almost 20% in this simulation, all six sets of well-chosen starting parameters resulted in approximately correct values for the iterative approach. The deviations in the estimated sensitivities reached up to 8%, whereas they were at most 2% for the specificities and the prevalence (see **Supplementary Table 7**).

The Bayesian method resulted in similar values for most of the well-chosen starting values with deviations up to 6% for the sensitivities and 5% for the specificities and the prevalence (see **Supplementary Table 19**). However, the model did not converge using starting value set 3, and the algorithm aborted the calculation.

Both the iterative, frequentist approach and the Bayesian approach resulted in strongly deviating parameter values with the poorly chosen starting values (see **Supplementary Tables 12, 24**).

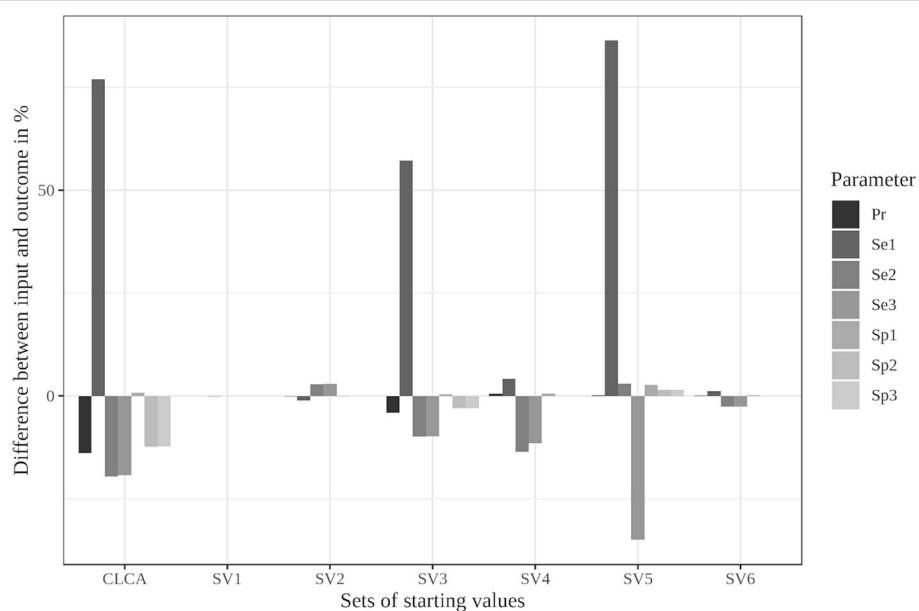


FIGURE 1 | Differences between the configured input data and model outcomes using six sets of starting values for the latent class model assuming conditional independence and the iterative model under the conditions of scenario 3 [CLCA, conditionally independent LCA; SV1, set of starting values 1 (Sv2-Sv6 are defined analogously); Pr, Prevalence; Se1, Sensitivity of test 1; Sp1, Specificity of test 1 (Se2, Se3, Sp2, and Sp3 are defined analogously)].

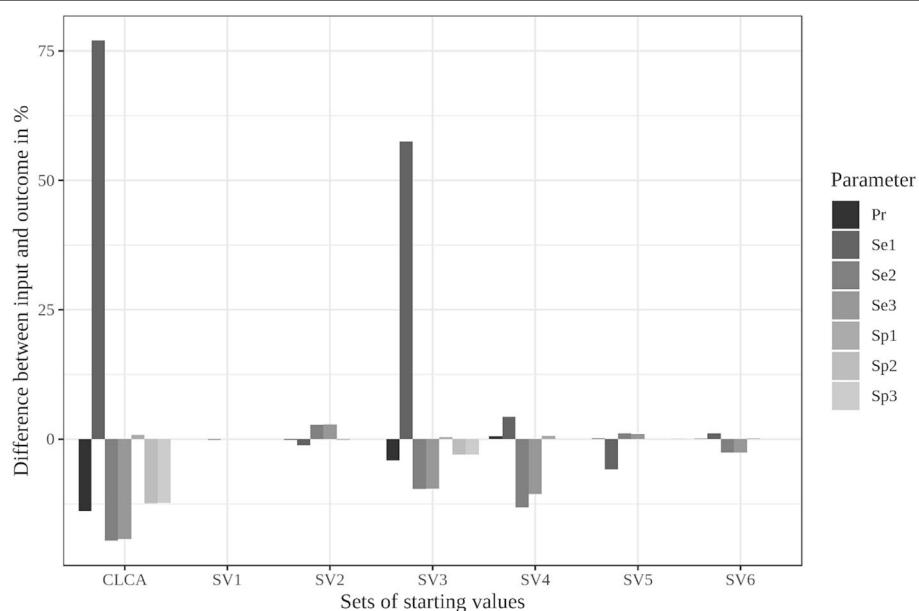


FIGURE 2 | Differences between the configured input data and model outcomes using the six sets of starting values for the latent class model assuming conditional independence and by using the iterative model under the conditions of scenario 3 with only positive pairwise dependencies allowed [CLCA, conditionally independent LCA; SV1, starting values 1 (Sv2-Sv6 are defined analogously); Pr, Prevalence; Se1, Sensitivity of test 1; Sp1, Specificity of test 1 (Se2, Se3, Sp2, and Sp3 are defined analogously)].

Parameter Restrictions

(Supplementary Tables 13, 14, 25, 26)

The limitation to reasonable positive dependencies had no effect on the results for most starting values since the resulting

dependencies were already within the defined boundaries. For starting value set 5, however, the differences between the input data and the outcome decreased remarkably (see Supplementary Table 13 and Figure 2). While it led to the

TABLE 2 | Maximum deviations of the three compared methods to the simulated values in the five simulated scenarios with well-chosen starting values displayed as values in percent.

	Independent LCA	Bayesian LCA	Iterative LCA
Scenario 1	1.0	6.9	4.4
Scenario 2	20.4	6.9	7.7
Scenario 3	77.0	15.0	86.4 ¹
Scenario 4	15.7	18.7 ²	5.9
Scenario 5	19.6	5.9 ³	8.1
Parameter restriction	77.0	15.1	57.5 ¹

¹Strongest deviation for the estimated sensitivity of test 1, the other parameters in the other scenarios had a maximum deviation of about 13%.

²Strongest deviation for the estimations of starting values 5, the parameters in the other scenarios had a maximum deviation of 11.8%.

³The model did not converge using starting values 3 and the algorithm aborted the calculation.

strongest deviations in the unrestricted model, the restricted model was able to reduce the differences to a maximum of 6%. Even the poorly chosen starting values reached partially better results when the boundaries were considered (see **Supplementary Table 14**).

In contrast, the parameter restrictions had very little effect in the Bayesian model (see **Supplementary Tables 25, 26**). Most results did not change at all. Some values, especially the prevalence estimates, were slightly worse (0.5%) in the restricted model.

Overall Results

The latent class analysis considering the dependency structure was able to calculate less biased parameter values than the classical frequentist latent class analysis for most of the informative starting values. Both the Bayesian and iterative methods produced very similar results (see **Table 2**). The estimated values differed considerably from the simulation settings if the prevalence of the simulated disease was low or if the starting values differed substantially from the simulated data. However, there were differences between the two methods. While the proposed iterative approach needed relatively accurate prior knowledge of the dependency structure, it was better able to deal with a slight-to-moderate initial misjudgment of the test accuracies and the prevalence than the Bayesian approach. However, in some rare situations, both methods are prone to restricted convergence or strongly biased results.

Although the iterative approach yielded varying results for the different starting value sets, the associated log-likelihood always had the same value within all five simulation scenarios. This finding indicates that all results within a scenario, although they have very different values, represent a local maximum and are therefore equally likely under the observed responses. These results were estimated in only a few iteration steps in all five scenarios (a maximum of 9 iteration steps in scenario 3, see **Supplementary Tables 5, 10**). Thus, the convergence of the iterative approach was not noticeably slower than that of the conditionally independent latent class model. However, it was significantly faster than the convergence of the Bayesian method, for which 11,000 iterations were set.

DISCUSSION

In this publication, we presented an iterative, frequentist latent class approach for the evaluation of conditionally dependent diagnostic tests. We compared it to the Bayesian method and the classical conditionally independent analysis by performing a simulation study.

If two diagnostic tests with the same biological principle are used, the same reasons (e.g., cross-reactions, pathogen concentration) will lead to incorrect diagnoses, which strongly connect the outcome of these tests (28). Applying the classical latent class model without consideration of these dependencies leads to results that differ strongly from the true underlying parameters. The simulation studies presented in this publication confirmed this hypothesis.

These differences decrease or are even eliminated by using a model that considers the conditional dependencies. However, the accuracy of the estimates from the presented iterative method as well as the Bayesian approach strongly depends on two factors: the starting values and the size of the underlying parameters.

Starting values for the test accuracy of the diagnostic tests used can be obtained from the manufacturer's evaluation studies or from previous studies employing these tests. The conditional dependencies can be estimated by examining the biological methods of the tests and comparing them to each other. Similar procedures are more likely to be highly dependent (21). However, if there is no prior knowledge for the applied tests and no information on the biological procedures used, it is difficult to determine the correct starting parameters, and they must be chosen randomly. Thus, there is no assurance that the selected parameters are actually correct. However, if good estimates are available, the presented iterative procedure will be able to find the right parameters in just a few steps.

The size of the underlying parameters also influences the quality of the estimates. For scenario 3 with a low prevalence and a strong dependency, all the compared methods attained suboptimal performance within the simulation study. This deteriorated accuracy in populations with low prevalence was also observed in the Bayesian framework in simulation studies (29). This finding could cause problems in applications since there are many diseases with low prevalence in veterinary medicine [e.g., (30, 31)]. However, there were differences in the results of the two approaches due to the different estimation methods. The iterative approach estimated the specificities of the presented simulation scenarios, especially in the third scenario, more precisely than the Bayesian method. Due to the higher values of the specificity, many consistently negative test results were correctly assigned to the negative latent class in the first step of the algorithm regardless of the starting values. The incorrectly assigned positive observations had little influence due to the low prevalence. Thus, higher test accuracies lead to a clearer separation of the latent classes and a greater tolerance toward poorly chosen starting values. Furthermore, the more the starting values deviated from the given settings, the greater the deviations in the results were. While the presented iterative approach fared slightly better with an initial misjudgment of the test accuracies and the prevalence, the Bayesian method led to a more accurate convergence when incorrect dependencies

were initially assumed. Therefore, the best latent class method for conditionally dependent diagnostic tests depends on the study population, the tests used and the accuracy of the existing prior knowledge.

Despite this phenomenon, the log-likelihoods of the different results within each scenario, the correct and the incorrect ones, converge to the same value. This finding suggests that the function has several maxima. Each result found by our method represents one local maximum and all these maxima are equally probable with the given dataset. Therefore, this model is not able to find a unique solution and well-chosen starting values are needed to ensure convergence to the correct parameter values. The reason is that the addition of dependency terms increases the number of parameters to be estimated, while the information provided by the observed response pattern does not change. The proposed method takes this property into account by estimating parameters in a stepwise algorithm that regards the dependency terms and the test accuracies alternately as fixed. As a result, there is a positive number of degrees of freedom in each step of the algorithm, and the identifiability is improved. The model is applicable to situations in which results from at least two diagnostic tests are available. However, as the two-test case was already underidentified without the additional dependency terms ($df = -1$) and therefore had no chance for a unique solution in the proposed iterative approach, we focused our analyses on the special case of three diagnostic tests. However, the increased number of tests is not sufficient for a clear result.

The ambiguity of the solutions occurs regardless of the chosen method, as the addition of dependency terms results in more parameters to be estimated than information is available in the data. Therefore, there is no model that takes the dependency between all tests into account and comes to a unique solution. This lack of knowledge has to be replaced by accurate prior information. If the priors are (unknowingly) false, the model is not always able to find the right solution for some parameter compositions. This limitation causes uncertainty regarding the assumptions at the beginning of the analysis; i.e., the results are not reliable, and not even an iterative calculation is able to solve this problem. Thus, good prior knowledge is necessary for accurate estimates, and uninformative priors should not be used with this method (the Bayesian or the iterative, frequentist approach). Other researchers have already observed and pointed out the importance of justified priors in the Bayesian framework (32, 33). Nevertheless, the Bayesian approach is able to express the certainty in the prior knowledge in the form of distribution functions, which may help to reduce the impact of initial misjudgments but also makes the modeling more complex, while the iterative, frequentist approach may include more “practical-use knowledge,” i.e., information on the general biological framework of the diagnostic method used.

Establishing boundaries for the dependencies improves the parameter estimates from the iterative approach further if the dependencies are within a certain interval and the values outside the interval can be excluded with certainty. These boundaries can prevent major deviations in the results, as shown in the second calculation of scenario 3. However, parameter restrictions

should be used with caution. Only unrealistic values for a certain application should actually be excluded. If that is not the case, a true underlying parameter value may unknowingly be rejected as a possible solution, and the algorithm is no longer able to calculate the correct parameter set. Hence, the restrictions help to improve parameter estimation but also bear the risk of excluding the correct results from the start by choosing incorrect limits. Thus, if one is unsure of which limits to choose, it is better to completely remove them and carry out the estimation only with the best possible starting values.

Overall, the fit of the latent class model and the parameter estimates can be improved by allowing an interaction term. If the results of three diagnostic tests are available, both the Bayesian method and the iterative, frequentist approach presented in this paper are strongly dependent on the prior information due to the lack of information in the data. If there is insufficient knowledge about the test accuracies, the prevalence and the dependencies of the tests, and hence, these values are initially misjudged, both methods will lead to incorrect results. Extensive prior knowledge is therefore the basis for the applicability of the latent class analysis considering of conditional dependencies, both in the Bayesian and frequentist frameworks.

CONCLUSION

The presented simulation study showed that considering a possible dependency structure improves the estimation in a latent class analysis. However, it was unable to clearly determine which method resulted in more accurate values overall, as the iterative, frequentist approach and the Bayesian approach performed differently in the presented scenarios. While both methods are dependent on prior knowledge in the form of well-chosen starting values and prior distributions, the simulation studies carried out in this publication suggest that the iterative, frequentist method requires previous knowledge that is oriented more toward practical experience and therefore may be easier to obtain.

Overall, the simulation studies presented here indicate that the iterative, frequentist approach is an appropriate method to evaluate conditionally dependent diagnostic tests.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

CS and AC: conceptualization. CS: data curation, formal analysis, investigation, methodology, software, validation, and writing—original draft. LK and AC: project administration and supervision. CS, LK, and AC: writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work is part of the Q fever - GermAn Interdisciplinary Program for reSearch and was funded by the Federal Ministry of Education and Research (BMBF) under project number 01KI1726B as part of the Research Network on Zoonotic Infectious Diseases. This Publication was supported by Deutsche Forschungsgemeinschaft and University of Veterinary Medicine

Hannover, Foundation within the funding programme Open Access Publishing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2021.588176/full#supplementary-material>

REFERENCES

- Dohoo IR, Martin W, Stryhn H. *Veterinary Epidemiologic Research*. 2nd ed. Prince Edward Island: AVC Incorporated Charlottetown (2009).
- Campe A, Abernethy D, Menzies F, Greiner M. Latent class regression models for simultaneously estimating test accuracy, true prevalence and risk factors for *Brucella abortus*. *Epidemiol Infect.* (2016) 144:1845–56. doi: 10.1017/S0950268816000157
- Gardner IA, Colling A, Greiner M. Design, statistical analysis and reporting standards for test accuracy studies for infectious diseases in animals: progress, challenges and recommendations. *Prev Vet Med.* (2019) 162:46–55. doi: 10.1016/j.prevetmed.2018.10.023
- Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *J Clin Epidemiol.* (1988) 41:923–37. doi: 10.1016/0895-4356(88)90110-2
- Branscum A, Gardner I, Johnson W. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev Vet Med.* (2005) 68:145–63. doi: 10.1016/j.prevetmed.2004.12.005
- Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics.* (1980) 36:167–71. doi: 10.2307/2530508
- Pouillot R, Gerbier G, Gardner IA. “TAGS”, a program for the evaluation of test accuracy in the absence of a gold standard. *Prev Vet Med.* (2002) 53:67–81. doi: 10.1016/S0167-5877(01)00272-0
- Vacek PM. The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics.* (1985) 41:959–68. doi: 10.2307/2530967
- Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics.* (2001) 57:158–67. doi: 10.1111/j.0006-341X.2001.00158.x
- Georgiadis MP, Johnson WO, Gardner IA, Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *J R Stat Soc Ser C.* (2003) 52:63–76. doi: 10.1111/1467-9876.00389
- Black MA, Craig BA. Estimating disease prevalence in the absence of a gold standard. *Stat Med.* (2002) 21:2653–69. doi: 10.1002/sim.1178
- Yang I, Becker MP. Latent variable modeling of diagnostic accuracy. *Biometrics.* (1997) 53:948–58. doi: 10.2307/2533555
- Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics.* (1996) 52:797–810. doi: 10.2307/2533043
- Albert PS, McShane LM, Shih JH, Network USNCIBTM. Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. *Biometrics.* (2001) 57:610–9. doi: 10.1111/j.0006-341X.2001.00610.x
- Zhang BC, Chen Z, Albert PS. Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts. *Biometrics.* (2012) 68:1294–302. doi: 10.1111/j.1541-0420.2012.01789.x
- Liu W, Zhang B, Zhang Z, Chen B, Zhou XH. A pseudo-likelihood approach for estimating diagnostic accuracy of multiple binary medical tests. *Comput Stat Data Anal.* (2015) 84:85–98. doi: 10.1016/j.csda.2014.11.006
- Hanson TE, Johnson WO, Gardner IA. Log-linear and logistic modeling of dependence among diagnostic tests. *Prev Vet Med.* (2000) 45:123–37. doi: 10.1016/S0167-5877(00)00120-3
- Politser P. Reliability, decision rules, and the value of repeated tests. *Med Decis Making.* (1982) 2:47–69. doi: 10.1177/0272989X8200200108
- Chieccchio A, Malvano R, Giglioli F, Bo A. Performance assessment of coupled tests: the effects of statistical non-independence. *Clin Chem Lab Med.* (1994) 32:169–76. doi: 10.1515/cclm.1994.32.3.169
- Toft N, Åkerstedt J, Tharaldsen J, Hopp P. Evaluation of three serological tests for diagnosis of Maedi-Visna virus infection using latent class analysis. *Vet Microbiol.* (2007) 120:77–86. doi: 10.1016/j.vetmic.2006.10.025
- Gardner IA, Stryhn H, Lind P, Collins MT. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev Vet Med.* (2000) 45:107–22. doi: 10.1016/S0167-5877(00)00119-7
- Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am J Epidemiol.* (1995) 141:263–72. doi: 10.1093/oxfordjournals.aje.a117428
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2019).
- Virella G. *Medical immunology*. 6th ed. New York, NY: Informa plc (2007).
- McKenzie BA. Veterinary clinical decision-making: cognitive biases, external constraints, and strategies for improvement. *J Am Vet Med Assoc.* (2014) 244:271–6. doi: 10.2460/javma.244.3.271
- Plummer M. *rjags: Bayesian Graphical Models using MCMC*. Vienna: R Foundation for Statistical Computing (2019).
- Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Stat Med.* (2009) 28:3049–67. doi: 10.1002/sim.3680
- Brenner H. How independent are multiple ‘independent’ diagnostic classifications? *Stat Med.* (1996) 15:1377–86. doi: 10.1002/(SICI)1097-0258(19960715)15:13<1377::AID-SIM275>3.0.CO;2-#
- McAlloon CG, Doherty ML, Whyte P, Verdugo C, Toft N, More SJ, et al. Low accuracy of Bayesian latent class analysis for estimation of herd-level true prevalence under certain disease characteristics—An analysis using simulated data. *Prev Vet Med.* (2019) 162:117–25. doi: 10.1016/j.prevetmed.2018.11.014
- Coulibaly ND, Yameogo KR. Prevalence and control of zoonotic diseases: collaboration between public health workers and veterinarians in Burkina Faso. *Acta Trop.* (2000) 76:53–7. doi: 10.1016/S0001-706X(00)00090-5
- Berger-Schoch A, Herrmann D, Schares G, Müller N, Bernet D, Gottstein B, et al. Prevalence and genotypes of *Toxoplasma gondii* in feline faeces (oocysts) and meat from sheep, cattle and pigs in Switzerland. *Vet Parasitol.* (2011) 177:290–7. doi: 10.1016/j.vetpar.2010.11.046
- Johnson WO, Jones G, Gardner IA. Gold standards are out and Bayes is in: implementing the cure for imperfect reference tests in diagnostic accuracy studies. *Prev Vet Med.* (2019) 167:113–27. doi: 10.1016/j.prevetmed.2019.01.010
- Williams CJ, Moffitt CM. Estimation of fish and wildlife disease prevalence from imperfect diagnostic tests on pooled samples with varying pool sizes. *Ecol Inform.* (2010) 5:273–80. doi: 10.1016/j.ecoinf.2010.04.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Schoneberg, Kreienbrock and Campe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

2.2 **Publikation II: An intercomparison study of ELISAs for the detection of porcine reproductive and respiratory syndrome virus – Evaluating six conditionally dependent tests**

Schoneberg C, Böttcher J, Janowetz B, Rostalski A, Kreienbrock L, Campe A (2022) An intercomparison study of ELISAs for the detection of porcine reproductive and respiratory syndrome virus – evaluating six conditionally dependent tests. PLoS ONE 17(1): e0262944. <https://doi.org/10.1371/journal.pone.0262944>

Journal	PLoS ONE
Editor	Brecht Devleesschauwer (Sciensano, Belgien)
Eingereicht	10. September 2021
Akzeptiert	08. Januar 2022
Veröffentlicht	25. Januar 2022

RESEARCH ARTICLE

An intercomparison study of ELISAs for the detection of porcine reproductive and respiratory syndrome virus – evaluating six conditionally dependent tests

Clara Schoneberg^{1*}, Jens Böttcher², Britta Janowetz², Anja Rostalski², Lothar Kreienbrock¹, Amely Campe¹

1 Department of Biometry, Epidemiology and Information Processing, WHO Collaborating Centre for Research and Training for Health in the Human-Animal-Environment Interface, University for Veterinary Medicine Hannover, Hannover, Germany, **2** Bavarian Animal Health Service, Poing, Germany

* clara.schoneberg@tiho-hannover.de



OPEN ACCESS

Citation: Schoneberg C, Böttcher J, Janowetz B, Rostalski A, Kreienbrock L, Campe A (2022) An intercomparison study of ELISAs for the detection of porcine reproductive and respiratory syndrome virus – evaluating six conditionally dependent tests. PLoS ONE 17(1): e0262944. <https://doi.org/10.1371/journal.pone.0262944>

Editor: Brecht Devleesschauwer, Sciensano, BELGIUM

Received: September 10, 2021

Accepted: January 8, 2022

Published: January 25, 2022

Copyright: © 2022 Schoneberg et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: The laboratory work was performed at the Bavarian Animal Health Service financially supported by the Free State of Bavaria and the Bavarian Joint Founding Scheme for the Control and Eradication of contagious Livestock Diseases (Bayerische Tierseuchenkasse). Publication was supported by Deutsche Forschungsgemeinschaft

Abstract

Latent class analysis is a widely used statistical method for evaluating diagnostic tests without any gold standard. It requires the results of at least two tests applied to the same individuals. Based on the resulting response patterns, the method estimates the test accuracy and the unknown disease status for all individuals in the sample. An important assumption is the conditional independence of the tests. If tests with the same biological principle are used, the assumption is not fulfilled, which may lead to biased results. In a recent publication, we developed a method that considers the dependencies in the latent class model and estimates all parameters using frequentist methods. Here, we evaluate the practicability of the method by applying it to the results of six ELISA tests for antibodies against the porcine reproductive and respiratory syndrome (PRRS) virus in pigs that generally follow the same biological principle. First, we present different methods of identifying suitable starting values for the algorithm and apply these to the dataset and a vaccinated subgroup. We present the calculated values of the test accuracies, the estimated proportion of antibody-positive animals and the dependency structure for both datasets. Different starting values led to matching results for the entire dataset. For the vaccinated subgroup, the results were more dependent on the selected starting values. All six ELISA tests are well suited to detect antibodies against PRRS virus, whereas none of the tests had the best values for sensitivity and specificity simultaneously. The results thus show that the method used is able to determine the parameter values of conditionally dependent tests with suitable starting values. The choice of test should be based on the general fit-for-purpose concept and the population under study.

and University of Veterinary Medicine Hannover, Foundation within the funding programme Open Access Publishing. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Porcine reproductive and respiratory syndrome (PRRS) is a disease in pigs caused by the *Betaarterivirus suis* (PRRSV). Two genotypes (1 and 2) are generally distinguished. Genotype 1 originated from Europe, and genotype 2 originated from North America; viruses within genotypes are not antigenically homogenous [1]. Infection is associated with late-term abortion in sows and respiratory disease in weaned and fattening pigs and thus results in significant economic losses worldwide [2]. Vaccination is frequently practiced to prevent clinical disease; currently, one genotype 2 and four genotype 1 live attenuated vaccines are commercially available in Germany.

After the discovery of the viral etiology of PRRS in the early nineties of the last century, an immune peroxidase monolayer assay (IPMA) was developed to detect antibodies [3]. Viral propagation in cell culture was extremely difficult. Initially, PRRSV was multiplied in primary lung alveolar macrophages, and later, the virus adapted to a permanent cell line (MARC145) [4]. Although the neutralization test is frequently regarded as a reliable gold standard for the detection of antibodies against viruses, in the case of PRRSV, neutralizing antibodies are developed only late in the course of infection and are quite specific to viral subtypes within genotypes [5, 6]. In contrast, IPMA allowed the detection of cross-reactive antibodies directed to the conserved and abundantly present nucleocapsid antigen. Finally, as IPMA was difficult to standardize, a patent-protected ELISA was developed and commercialized (see Test 1). Since then, a huge body of experience has accumulated, and this ELISA has been established in worldwide laboratories. After the patent expired, additional ELISAs were developed and required validation. The question for a proper gold standard thus re-emerged. At least for some authors, it became a matter of course to regard the existing ELISA as a gold standard [7]. However, this view precludes any further improvement of antibody ELISAs; therefore, an alternative approach is required.

In a voluntary study by the Bavarian Animal Health Service, six of these ELISAs were used on sows, gilts and piglets from farms in southeast Germany with different vaccination statuses. The aim of the study was to estimate the accuracy of the tests and to assess the applicability of the newly developed ELISAs both under in-field conditions.

Generally, diagnostic test evaluation is performed by comparison to a gold standard test [8]. If there is no gold standard or if it only has suboptimal accuracies, a comparison may lead to incorrect values [9]. The application of the tests to a mostly small group of animals with a known disease status leads to results that only apply to the observed population. Therefore, latent class analysis (LCA) is often used as a method for estimating the diagnostic test accuracy for individuals with unknown disease status without a gold standard [10].

The LCA assumes that an underlying latent structure exists and that the diagnostic tests that measure the true unknown disease status of the individuals represent imperfect indicator variables under investigation. Based on the response patterns of the tests, both the accuracy of the tests and the proportion of both disease statuses (positive/negative) are discovered by the model. Hence, the test performance under the given conditions, such as the structure of the population in terms of age, gender, race, immune status, etc., can be estimated. The parameters may be estimated using Bayesian [11] or frequentist methods [12, 13].

A requirement of LCA is the conditional independence of the diagnostic tests [14]. This means that within a latent class, the result of one test does not allow conclusions to be drawn about the results of the other tests. However, if infectious diseases are considered, this condition is often not met. Conditionally independent tests use different biological principles (e.g., detection of antigen or antibodies), which cannot always be detected at the same time of infection [15]. Thus, these tests do not measure the same latent status, and applying the LCA leads

to biased results. Therefore, LCA requires tests with the same biological basis, which means that they are conditionally dependent. This is the case in the study for the detection of PRRSV. All tests used assess the presence of antibodies against the virus, i.e., they measure the same latent status.

There are already some approaches that consider conditional dependencies both in the Bayesian [16–18] and frequentist [19–24] frameworks. These methods often require some assumptions, such as a large number of diagnostic tests or accurate prior distributions. We developed a frequentist latent class approach, which takes the dependencies of the tests into account, thus no longer requiring the assumption of conditional independence [25]. In addition, this method is easily applicable in many situations even when little prior information is accessible. This approach expands the LCA by a term for the dependency and estimates the parameters alternately in an iterative process until convergence occurs. We showed in simulation studies that this method is better than the classic conditionally independent LCA in many situations and provides results similar to the Bayesian approach for conditionally dependent tests; however, in contrast this approach does not need any prior distribution.

In this publication, we show that our iterative, frequentist latent class approach is applicable to real-world data by evaluating the results of the mentioned voluntary study by the Bavarian Animal Health Service containing the results of six different ELISA tests against the PRRS virus for pigs in southern Germany. We present the results of the analysis of the entire dataset as well as a vaccinated subgroup of animals. Different approaches for the starting values of the algorithm are considered, and the resulting parameter values are compared.

Material and methods

Sample collection and diagnostics applied

Between 2016 and 2018, the Bavarian Animal Health Service offered voluntary PRRSV immune control to pig farmers in southeast Germany. Therefore, per farm, blood samples were collected from ten weaned piglets at 8–12 weeks, six gilts and 18 sows of different parities (1./2., 3./4. and 5./6. parity). After initial testing (PCR, ELISA Test 1, neutralization tests against and IFN-recall assay after stimulation of whole blood with a panel of PRRSV-vaccine virus strains), sera were stored at -20°C. The current vaccination status of the herd was assessed by a face-to-face questionnaire. Twenty-four farms in 13 districts participated. In 2019, sera were tested for antibodies against the virus by six different ELISA tests from five manufacturers:

- PRRS X3, Idexx (Test 1)
- pigtyp PRRSV Ab, Indical (Test 2)
- ID Screen PRRS indirect, IDVet (Test 3)
- Ingezim PRRS 2.0, Ingenasa (Test 4)
- Ingezim PRRS universal, Ingenasa (Test 5)
- PrioCHECK, ThermoFisher (Test 6).

A single batch was used for each ELISA, and all sera per test were analyzed in one run. The farms had been immunized in different ways in the past. Fourteen farms received vaccinations against genotype 1 of the virus, 4 farms received vaccinations against genotype 2 of the virus, 1 farm was vaccinated against both genotypes, and 5 farms did not have any vaccination in the past. To ensure comparability of the test results, only tests that checked for both genotypes of the virus were used.

In this analysis data from 812 pigs were evaluated. Overall, 8 animals with incomplete results were documented; therefore, they were removed from further analysis. Thus, 804 animals were included in the statistical evaluation.

The analysis of the dataset determined the test accuracies and the prevalence in a sample selected from the entire population without specific selection criteria. The sample consisted of animals with different vaccination statuses: unvaccinated animals and animals vaccinated against genotype 1, genotype 2 or both genotypes of the virus. Thus, great heterogeneity in the immune status of the animals is present as a typical sample from pig farms in Germany. To examine whether this heterogeneity impacts the performance of the method and how the test accuracy and prevalence change, we applied the methods used to a uniformly vaccinated population for an additional analysis. To examine a group as homogeneous as possible with a large number of cases, we considered the 468 samples on the 14 farms vaccinated against European genotype 1 of the virus for the second part of the analysis.

The latent class model

The intercomparison of diagnostic tests was performed with the method introduced recently using a modified latent class model [25]. In a latent class analysis, it is assumed that there is a latent variable with two classes. When applied to diagnostic tests, these classes (diseased vs. not diseased) are defined by the observed diagnostic tests, which assume two values (positive/negative). Each individual tested by all diagnostic tests analyzed in the LCA represents an observation in the model. Thus, the proportion of the positive class, i.e., the prevalence, and the response probability of each test in these classes (i.e., the test accuracies) can be determined for the sample. Our model also considers conditional dependencies between the tests due to similar test principles by integrating an additional dependency term into the model.

Formally, in the modified model, it is assumed that there is a latent variable with two classes, which is measured by M observed variables. Observation $Y_i = (Y_{i1}, \dots, Y_{iM})$, $i = 1, \dots, N$ represents individual i 's response pattern with the possible values $r_m = 0, 1$ for observation Y_{im} . The probability of membership in latent class c can be expressed as γ_c with $\sum_{c=0}^1 \gamma_c = 1$, and the probability of the response r_m to variable m in class c can be expressed as $\rho_{m,r_m|c}$. Let $I(\cdot)$ be the indicator function. Then, the likelihood of parameters γ and ρ for the observations Y can be written in the following form:

$$L(\gamma, \rho | Y) = \sum_{i=1}^N \sum_{c=0}^1 \gamma_c \left(\prod_{m=1}^M \prod_{r_m=0}^1 \rho_{m,r_m|c}^{I(Y_{im}=r_m)} + \eta_{r_1, \dots, r_M}^c \right), \quad (a)$$

where η_{r_1, \dots, r_M}^c is a term for the influence of the $\sum_{i=2}^M \binom{M}{i}$ conditional two-way to M -way

dependencies of all M tests in class c on the likelihood of the respective observed response patterns as described in [25]. The conditional dependencies are expressed in terms of the conditional covariances between the test results in the two latent classes, respectively. They are considered as the differences between the observed proportion of matching correct responses of two to M tests in latent class c and the expected corresponding proportion in class c under the conditional independence assumption.

All parameter values from Formula (a) are estimated in an iterative algorithm:

1. Appropriate starting values are chosen for the prevalence, test accuracies and dependencies.
2. The dependencies are considered fixed. The test accuracies and the prevalence are newly estimated by an expectation maximization (EM) algorithm in the same manner as noted in the conditionally independent LCA (such as in [13]).

3. Using the test accuracies and the prevalence estimated in the previous step, the expected frequencies in the case of conditional independence are redetermined. The dependencies can then be recalculated from the difference to the observed frequencies.
4. The algorithm starts again with step 2 until it converges, i.e., the difference of the log-likelihood of two successive steps falls below a specified limit or the algorithm reaches the maximum number of 1000 iterations.

Confidence intervals for test accuracies and the prevalence are calculated using the normal approximation interval for binomial distributions. The model is applicable if the results of at least three tests, each with two possible answers (diseased / not diseased), are available. The procedure was implemented in R (version 3.5.0; [26]); see [S1 File](#).

Starting values

In a previous investigation, we showed that the applied iterative frequentist approach of the LCA needs well-chosen starting values to converge to the true underlying values [25]. Therefore, we used three sources of information for the parameters required: the manufacturers' evaluation studies conducted for every commercially available test, previous publications evaluating the diagnostic tests used in this study [7, 27–34] and experience from application from the researchers of the Bavarian Animal Health Service. Additionally, we used the examined dataset to estimate the prevalence in the sample.

Based on these sources, we proceeded as follows to determine the starting values that can be found in [S1](#) and [S3](#) Tables (see [Fig 1](#) also):

1. Determine test accuracies based on each available source:
 - a. Adopt values from the manufacturer's evaluation studies (indicated as M).
 - b. Find previous publications (indicated as P) evaluating the diagnostic tests used in this study (see [S5 Table](#)) and calculate a weighted average by considering the studies that have a similar structure and study population more strongly.
 - c. Ask researchers (indicated as R) from the associated laboratories (here, the Bavarian Animal Health Service) for an assessment of the test accuracies based on their experience in application and the values of the other sources.
2. Estimate the dependencies of the tests in terms of their correlation (as described in [25]) based on each available source, so that these values are standardized and independent of the test accuracies:
 - a. Find previous publications evaluating the diagnostic tests used in this study and use their values for conditional dependency (here, we could not find any previous publications considering a dependency of the tests, so this source did not provide any information on the conditional dependency of the tests).
 - b. Ask researchers (indicated as R) from the associated laboratories (here, the Bavarian Animal Health Service) for an assessment of the dependencies of the tests used based on their experience and the biological principles of the tests used.
3. Establish prevalence for the dataset under consideration by estimating the approximate proportion of positive tests in the dataset, considering the chosen starting values of the test accuracies from step 1 (here, using the dataset ([Table 1](#)) with a slightly higher prevalence, as all sources in step 1 assume a higher sensitivity than specificity for all six tests).

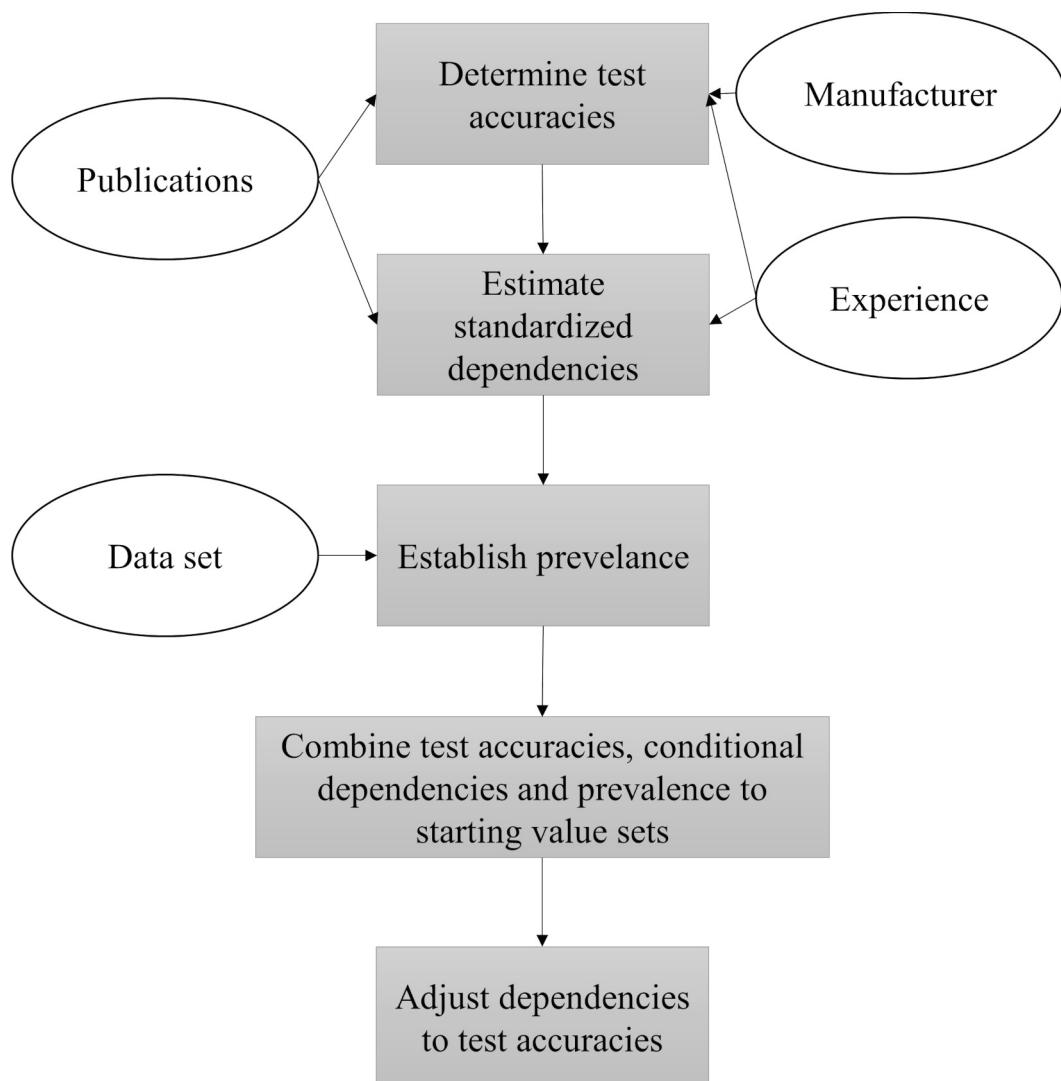


Fig 1. Step-by-step procedure for determining the starting values when taking various information sources into account.

<https://doi.org/10.1371/journal.pone.0262944.g001>

4. Define starting value datasets: All available information on the test accuracies, the dependencies and the prevalence are combined. Every possible combination forms a starting value set. The exact number of the resulting sets depends on the amount of prior

Table 1. Observed frequencies of the two results for all six tests used in the latent class analysis of the complete dataset (percentage rounded to one digit).

	Positive tested Animals (in %)	Negative tested Animals (in %)
Test 1	549 (68.3)	255 (31.7)
Test 2	518 (64.4)	286 (35.6)
Test 3	487 (60.6)	317 (39.4)
Test 4	461 (57.3)	343 (42.7)
Test 5	578 (71.9)	226 (28.1)
Test 6	623 (77.5)	181 (22.5)

<https://doi.org/10.1371/journal.pone.0262944.t001>

information. At least four different starting values should be applied so that two different dependency structures and two different test accuracy sets are used to examine the stability of the results. If there is only one source for the dependencies, independence of the tests (indicated as I) can also be used as a starting value to test stability of the method.

Here, that means all three starting values for the test accuracies (M, P, R) are combined with the information for prevalence (step 3) and conditional dependencies (R). In addition, all three values for the test accuracies are used as starting values with the assumption of conditional independence (I). This leads to six different starting value sets (MI, MR, PI, PR, RI, RR, see [S1 Table](#)).

5. Adjust the standardized values of the dependencies to the selected test accuracies M, P, R and the prevalence (step 3) so that the conditional dependencies are described in terms of conditional covariances and any combination of the test results has a positive probability (see [25]). This leads to deviating values of the conditional covariances depending on the selected starting values for the test accuracies.

The above-described steps for the starting values 1a and 1c were performed a second time for the vaccinated subgroup. Since this is a subgroup of the first dataset, we assumed that it has parameters similar to those of the entire dataset. Therefore, we selected the starting values of the test accuracies that differ the most from one another to assess the stability of the model. All starting values for the vaccinated subgroup can be found in [S2 Table](#).

Results

In the following section, we present three main results of the analysis. First, we show the parameter values calculated by the method for the complete dataset. Second, these values are considered in context with the selected starting values. Finally, the results are compared with those of the vaccinated subgroup to obtain an impression of the influence a vaccination has on the performance of the method proposed and the resulting parameter values. The starting value sets MI, MR, PI, PR, RI, RR specified in the last chapter are used in the calculation, where M indicates the information provided by the manufacturer, R denotes the information on test accuracies and dependencies provided by the researchers, P indicates the values from previous publications and I represents the assumptions of independent tests.

Overall results and effect of starting values

In general, all calculations agreed that slightly greater than three-quarters (75.3–76.8%) of the animals in the sample examined were assigned to the positive latent class, i.e., had antibodies against the virus. Test 6 (PrioCHECK) had the highest estimated sensitivity values, and test 4 (Ingezim PRRS 2.0) had the lowest estimated sensitivity values (see [Table 2](#)). The values of the remaining tests were distributed evenly between these results. Overall, the specificities had higher values, which were between 95% and 99% for tests 1 to 4. Tests 5 (Ingezim PRRS universal) and 6 had lower specificities of approximately 88% and 78%, respectively. Test 6 in particular produced an extended number of false-positive results. Test 1 (PRRS X3) had the highest values of sensitivity and specificity combined with minimal differences compared with the other tests considered. These results were consistent with the observed frequencies ([Table 1](#)).

In the negative latent class, tests 1, 2, 3 and 4 had strong pairwise dependencies. Tests 5 and 6 were less dependent on each other and on the other tests. This yields a greater proportion of the agreeing negative results in the first four tests due to simultaneous incorrect results than in

Table 2. Estimated values for the prevalence and the test accuracies for the six starting value sets for the complete sample with confidence limits reported in brackets (rounded to one digit).

Parameters estimated	Starting value sets					
	Set MI	Set MR	Set PI	Set PR	Set RI	Set RR
Prevalence in %	75.5 [72.5, 78.4]	76.7 [73.7, 79.6]	75.7 [72.7, 78.6]	76.8 [73.9, 79.7]	75.3 [72.3, 78.3]	76.4 [73.4, 79.3]
Sensitivity in %						
Test 1	88.9 [86.7, 91.0]	87.7 [85.4, 89.9]	88.6 [86.5, 90.8]	87.6 [85.3, 89.8]	89.0 [86.8, 91.2]	87.9 [85.7, 90.2]
Test 2	84.4 [81.9, 86.9]	83.1 [80.5, 85.7]	84.2 [81.6, 86.7]	83.0 [80.4, 85.6]	84.5 [82.0, 87.0]	83.4 [80.8, 85.9]
Test 3	79.7 [76.9, 82.5]	78.5 [75.6, 81.3]	79.5 [76.7, 82.3]	78.3 [75.5, 81.2]	79.8 [77.1, 82.6]	78.7 [75.9, 81.5]
Test 4	75.6 [72.6, 78.6]	74.5 [71.5, 77.5]	75.4 [72.4, 78.4]	74.4 [71.4, 77.4]	75.8 [72.8, 78.7]	74.8 [71.8, 77.8]
Test 5	91.3 [89.3, 93.2]	90.5 [88.5, 92.5]	91.1 [89.2, 93.1]	90.4 [88.3, 92.4]	91.4 [89.5, 93.4]	90.7 [88.7, 92.7]
Test 6	95.3 [93.8, 96.7]	95.0 [93.5, 96.5]	95.1 [93.7, 96.6]	94.9 [93.4, 96.4]	95.3 [93.9, 96.8]	95.1 [93.7, 96.6]
Specificity in %						
Test 1	94.9 [93.4, 96.4]	95.4 [94.0, 96.9]	95.1 [93.6, 96.6]	95.5 [94.1, 97.0]	94.8 [93.3, 96.4]	95.2 [93.8, 96.7]
Test 2	96.8 [95.6, 98.1]	96.9 [95.7, 98.1]	96.9 [95.7, 98.1]	96.9 [95.8, 98.1]	96.8 [95.6, 98.0]	96.8 [95.5, 98.0]
Test 3	98.2 [97.3, 99.1]	98.2 [97.2, 99.1]	98.3 [97.4, 99.2]	98.2 [97.3, 99.1]	98.2 [97.2, 99.1]	98.1 [97.2, 99.0]
Test 4	98.8 [98.1, 99.6]	99.1 [98.4, 99.7]	98.9 [98.1, 99.6]	99.1 [98.4, 99.7]	98.8 [98.1, 99.6]	99.0 [98.3, 99.7]
Test 5	87.8 [85.5, 90.0]	89.2 [87.0, 91.3]	88.0 [85.7, 90.2]	89.2 [87.1, 91.4]	87.6 [85.3, 89.9]	88.9 [86.8, 91.1]
Test 6	77.1 [74.2, 80.0]	80.0 [77.2, 82.7]	77.4 [74.5, 80.3]	80.1 [77.4, 82.9]	77.0 [74.0, 80.0]	79.6 [76.8, 82.4]

<https://doi.org/10.1371/journal.pone.0262944.t002>

the other two tests. All pairwise dependencies of the tests were substantially higher within the positive latent class. Tests 1, 5 and 6 had the highest dependencies in the positive latent class, which were almost twice as large as the strongest dependencies in the negative class. Therefore, the dependency structures in the two classes differ both in the tests with the highest dependencies and in the size of these values.

The calculated dependencies of the tests for all six sets of starting values can be found in [S3 Table](#).

All six starting value sets led to very similar results in the test accuracies and the prevalence ([Table 2](#)). The estimates of prevalence and sensitivity exhibited maximum deviations of 1.5%. The results for the specificity had an even greater consistency. Only the estimations of the specificity of test 6 varied by up to 3%. Thus, all six runs of the algorithm yielded consistent results despite differing starting values in contrast to the first simulation study [[25](#)].

Vaccinated subgroup

As the evaluation of diagnostic tests is related to the population under study, we additionally compared the results with the subgroup of farms that had participated in a vaccination

programme before our investigation. The results in this vaccinated subgroup were less clear. Starting value set 1 led to a significantly reduced prevalence and lower values for all specificities (see [Table 3](#)). According to these results, the vaccinated subgroup included 68.7% antibody-positive animals. The specificities were between 89.8% for test 4 and 41.5% for test 6. On the other hand, the sensitivity was (almost) 100% for all six tests.

In contrast, the other three starting value sets led to a significant deviation from the calculated parameter values of starting value set RI. These results were very similar to one another ([Table 4](#)). In total, 83.4–88.0% of animals were antibody positive, which is approximately 10% greater than that in the complete dataset, whereas the calculated values of the sensitivities and specificities were similar to those of the entire dataset. However, noticeable deviations of up to 11.6% were noted between the estimated parameter values of the different starting values.

The dependencies of the tests were similar to those of the entire dataset. Only starting value set RI led to differing values. Thus, some dependency values were negative, which implies that one test has an increased probability of a positive result while that probability is reduced for the other test.

Discussion

In this publication, we applied an iterative, frequentist latent class approach to real-life data of six conditionally dependent ELISA tests for PRRSV on pig farms with different vaccination statuses.

Overall results

The analysis produced clear results for the entire dataset. All runs of the algorithm led to results with a maximum deviation of about three percentage points from the other results for all parameters under study, although the selected starting values differed strongly. Therefore, there seems to be only one maximum of the likelihood function, and these results can be considered reliable.

Table 3. Resulting values for the prevalence and the test accuracies for starting value set RI (test accuracies estimated by researchers with assumption of independent tests) for the complete sample and the vaccinated subgroup with confidence limits reported in brackets (rounded to one digit).

Starting value sets	Complete dataset	Vaccinated subgroup
Prevalence in %	75.3 [72.3, 78.3]	68.7 [64.6, 72.9]
Sensitivity in %		
Test 1	89.0 [86.8, 91.2]	100.0 [99.7, 100.0]
Test 2	84.5 [82.0, 87.0]	98.2 [97.0, 99.4]
Test 3	79.8 [77.1, 82.6]	97.4 [95.9, 98.8]
Test 4	75.8 [72.8, 78.7]	95.5 [93.6, 97.3]
Test 5	91.4 [89.5, 93.4]	100.0 [100.0, 0.0]
Test 6	95.3 [93.9, 96.8]	100.0 [100.0, 0.0]
Specificity in %		
Test 1	94.8 [93.3, 96.4]	71.0 [66.9, 75.1]
Test 2	96.8 [95.6, 98.0]	80.2 [76.6, 83.8]
Test 3	98.2 [97.2, 99.1]	87.9 [85.0, 90.9]
Test 4	98.8 [98.1, 99.6]	89.8 [87.1, 92.6]
Test 5	87.6 [85.3, 89.9]	59.4 [55.0, 63.9]
Test 6	77.0 [74.0, 80.0]	41.5 [37.0, 46.0]

<https://doi.org/10.1371/journal.pone.0262944.t003>

Table 4. Resulting values for the prevalence and the test accuracies for the four starting value sets for the part of the sample that is vaccinated against genotype 1 of PRRSV with confidence limits reported in brackets (rounded to one digit).

Starting value sets	Set MI	Set MR	Set RI	Set RR
Prevalence in %	83.4 [80.0, 86.8]	88.0 [85.0, 90.9]	68.7 [64.6, 72.9]	87.4 [84.4, 90.4]
Sensitivity in %				
Test 1	91.8 [89.3, 94.3]	87.9 [85.0, 90.9]	100.0 [99.7, 100.0]	88.4 [85.5, 91.3]
Test 2	87.6 [84.6, 90.6]	83.4 [80.0, 86.7]	98.2 [97.0, 99.4]	83.9 [80.6, 87.2]
Test 3	84.2 [80.9, 87.5]	80.1 [76.5, 83.7]	97.4 [95.9, 98.8]	80.6 [77.1, 84.2]
Test 4	82.2 [78.8, 85.7]	78.1 [74.3, 81.8]	95.5 [93.6, 97.3]	78.6 [74.9, 82.3]
Test 5	94.6 [92.5, 96.6]	91.3 [88.8, 93.9]	100.0 [100.0, 0.0]	91.8 [89.3, 94.3]
Test 6	99.0 [98.1, 99.9]	96.9 [95.3, 98.4]	100.0 [100.0, 0.0]	96.1 [94.3, 97.9]
Specificity in %				
Test 1	92.5 [90.1, 94.9]	96.3 [94.6, 98.0]	71.0 [66.9, 75.1]	95.5 [93.7, 97.4]
Test 2	96.1 [94.3, 97.8]	96.9 [95.3, 98.4]	80.2 [76.6, 83.8]	97.0 [95.4, 98.5]
Test 3	96.8 [95.2, 98.4]	98.0 [96.7, 99.2]	87.9 [85.0, 90.9]	98.0 [96.7, 99.3]
Test 4	98.8 [97.7, 99.8]	99.0 [98.1, 99.9]	89.8 [87.1, 92.6]	98.9 [97.9, 99.8]
Test 5	83.3 [79.9, 86.7]	89.3 [86.5, 92.1]	59.4 [55.0, 63.9]	88.7 [85.8, 91.5]
Test 6	72.1 [68.1, 76.2]	83.7 [80.3, 87.0]	41.5 [37.0, 46.0]	74.6 [70.7, 78.5]

<https://doi.org/10.1371/journal.pone.0262944.t004>

When interpreting the results, the methods of data collection must be considered. The samples were obtained during a voluntary examination that focused on the immune status after vaccination. As a result, only farms where the farmers had a high level of interest in the immune status of their animals and very few unvaccinated farms were included in the study. Therefore, this dataset is a convenience sample, and the proportion of unvaccinated farms is less than the proportion of unvaccinated farms in the total population, which is approximately 52% according to information from the Bavarian Animal Health Service. Thus, the calculated values for the test accuracies and the prevalence might not reflect the total population, as these values depend on the immune status of the animals examined [35]. However, this study provided a range of values of all the test accuracies indicating the relationship between the tests, which can serve as a basis for other investigations on the diagnostic tests used.

Estimation difficulties in the vaccinated subgroup

The results for the analysis of the subset of the farms vaccinated against the EU strains of the virus were less clear, as the resulting values differed depending on the selected starting value. In particular, the prevalence and PrioCHECK values varied up to 42%. This finding suggests that the information on if and to which extent a vaccination was applied in the population under study is crucial for the evaluation. All four runs of the algorithm provided results that have the same value of the likelihood function. Therefore, each calculated parameter combination represents the unknown true parameters with the same probability under the given dataset. Thus, there is no unique solution, and the composition of the study population and the biological principles of the tests used must also be considered for interpretation.

Although the results of starting value sets MI, MR and RR had a high level of agreement, the calculated parameters for starting value set RI deviated from these values. In comparison, starting value set RI led to a lower prevalence, partially negative dependencies and 100% sensitivities for three of the tests. Perfect tests (i.e., perfect sensitivities) are not to be expected for the applied ELISA tests. Although a changed sensitivity of the tests used due to the selection of an equally vaccinated subgroup with less variation in genotype was possible, technical errors

and an insufficient antibody concentration in the sample were still possible [36]. A reduced prevalence in a subgroup with more vaccinated animals is also very unlikely. Moreover, negative conditional dependencies of the tests imply that the probability of deviating results between two tests is increased compared to independent tests. Due to the same biological principle of the tests, this result is not plausible. Thus, the results of starting value sets MI, MR and RR seem to be more likely to represent the true parameter values. In that case, the proportion of animals with antibodies was 8–13% greater than that in the entire dataset. Therefore, vaccination against the European strain of the virus seems to have been successful, with a proportion of approximately 83–88% antibody-positive animals.

The reason that these problems exclusively occur in the vaccinated subgroup despite the same starting values may be the fact that the animals were all vaccinated against the same genotype of the virus. This feature causes a stronger homogeneity in the immune status of the animals and results in a homogenous dataset regarding positive test results in different diagnostic tests. This subsequently means that the statistical method has fewer deviating response patterns available for estimation, which increases the variation of the results from different starting values. As this was not predictable after the initial evaluation studies [25], it is not possible to indicate a percentage threshold of vaccinated individuals above which such problems can occur based on this study. This underlines the fact that prevaccination information is crucial for the evaluation.

The high proportion of vaccinated animals in this sample complicated the estimation for the statistical method used, but incorrect results can be discovered using several carefully selected starting values. If the results of all starting values agree (as with the entire dataset), then there is a clear indication that these are the true parameter values. If there are deviating results (as in the vaccinated subgroup), implausible results can be determined by considering the test principles used and the population examined. Hence, the method is also applicable in the case of a vaccinated population.

Effect of starting values

In the vaccinated subgroup differences in the results may appear, due to poorly chosen starting values. This was previously shown in simulation studies [25]. Therefore, we used starting values that we considered similar to the true parameter values based on various sources. However, these values differ only slightly and were in part selected based on subjective evaluations. Since there was a possibility that these assumptions were wrong, we systematically varied the individual parameter groups of the researchers' assessment (starting values RR) for the entire data set in a sensitivity analysis to check what effect changed starting values may have on the result ([S6 Table](#)). These analyses showed that deviating starting values lead to different results in some cases ([S7 Table](#)). For example, some changed starting values resulted in a prevalence of 0% or negative dependencies, which does not seem plausible under given circumstances. This underlines the conclusion made in the previous publication that the choice of starting values is important.

The assumption of independent tests seems to deviate greatly from the true parameter values, depending on the chosen initial values of test accuracies. Therefore, independent tests do not seem to be a good choice for starting values for this dataset. The result of starting value set RI, which assumed independent tests, deviates strongly from the other biologically more plausible results. Independence of the tests is an obvious and easy-to-implement choice for the starting values. However, in the case of several ELISA tests with the same biological principle, it would be a rather inapt assumption. Therefore, analyzing the biological test principles used combined with experience from application seems to be the more appropriate approach for

identifying starting values for the dependency, as that led to values that were more similar to the outcomes. However, starting value set MI, which also assumed independence, did not lead to implausible results. This result indicates that the selected initial test accuracies also influence the result. Some starting parameter constellations of test accuracies were closer to the results and therefore better suited as a starting point for the algorithm. Experience from application on the test accuracies led to starting values that were similar to the outcome. Previous publications also provided information on the values of the test accuracies. However, clear differences exist between the animals examined in the different studies in terms of species, sex and immune history. The statistical methods used also differed between the various publications. These factors influence the calculated values for the test accuracies [8, 35], which led to different results in the examined studies. These differences made determining the starting values based on the results of previous publications challenging. However, considering these influences, we were able to determine well-fitted starting values. The values from the manufacturers' studies were always greater than the values calculated in this study, which was probably also due to the different study situations.

Overall, the sources used for the starting values were useful for the application to our statistical method. Some values were closer to the results than others. Given that this difference is due to the composition of the population examined in this study, this may be different from other studies on other diseases.

Accuracy and applicability of the tests examined

Overall, the presented method calculated predominantly high accuracies for the diagnostic tests used in the complete dataset. We determined values of 74.4–95.3% for the sensitivity and 76.9–99.1% for the specificity of the tests. However, none of these tests had the best values for sensitivity and specificity simultaneously. Thus, the method again shows that there is no general gold standard that is applicable for all situations. The optimal test depends on the objective of the investigation and whether the false-negative or false-positive results are to be minimized. If the user wants to have a high level of certainty with the positive results and thus reduce the possibility of false-positive results, then a test with high specificity, such as the Ingezim PRRS 2.0, may represent the optimal test. Conversely, if all positive animals are actually to be identified as such, a test with maximum sensitivity (e.g., PrioCheck) might be the right choice. The accuracy of the diagnosis can also be further increased if a sequential test procedure is used [8]. For example, if false-positive animals are to be excluded with maximum certainty, it may be helpful to first sample the animals with a test with an overall very high accuracy (e.g., PRRS X3). All individuals with a positive test are then also sampled with a test with maximum specificity (e.g., ID Screen PRRS indirect or Ingezim PRRS 2.0), and only observations with two positive test results are then considered positive. This process minimizes the risk of a false-positive assessment and therefore an unnecessary intervention. However, the dependency of the tests must be taken into account when combining them. If two tests have a very high dependency, the same factors lead to incorrect results. A high dependency increases the probability of matching false results [37].

Conclusion

In this publication, we calculated the test accuracies of six ELISA tests used for the detection of PRRSV in pigs in southern Germany. Using the examined dataset as an example, we were able to demonstrate that the applied latent class approach is able to determine the parameter values of conditionally dependent tests with suitable starting values. Different methods of choosing starting values were shown.

None of the tests used had the best sensitivity and specificity simultaneously. Hence, the detection method must be chosen depending on the general fit-for-purpose concept and the sample population under study.

Here, the estimation of the parameters in a vaccinated subgroup was less precise, which suggests the need to take the heterogeneity and homogeneity of the immune status into account.

Supporting information

S1 Table. Starting values for the stepwise latent class algorithm for the complete data set.
(DOCX)

S2 Table. Starting values for the stepwise latent class algorithm for the vaccinated subgroup.
(DOCX)

S3 Table. Resulting values for the stepwise latent class algorithm for the complete data set for the five starting value sets.
(DOCX)

S4 Table. Resulting values for the stepwise latent class algorithm for the vaccinated subgroup for the three starting value sets.
(DOCX)

S5 Table. Studies evaluating the diagnostic test accuracies of the tests analyzed in this publication that were used as sources of information for suitable starting values for the iterative, frequentist latent class analysis.
(DOCX)

S6 Table. Starting values for the sensitivity analysis of the stepwise latent class algorithm.
(DOCX)

S7 Table. Resulting values of the stepwise latent class algorithm for the for the sensitivity analysis.
(DOCX)

S8 Table. Raw data of the study.
(XLSX)

S1 File. The R code of the algorithm.
(R)

Author Contributions

Conceptualization: Jens Böttcher, Amely Campe.

Data curation: Clara Schoneberg.

Formal analysis: Clara Schoneberg.

Funding acquisition: Jens Böttcher, Amely Campe.

Investigation: Clara Schoneberg, Jens Böttcher, Britta Janowetz, Anja Rostalski.

Methodology: Clara Schoneberg.

Project administration: Jens Böttcher, Lothar Kreienbrock, Amely Campe.

Resources: Jens Böttcher, Britta Janowetz, Anja Rostalski.

Software: Clara Schoneberg.

Supervision: Lothar Kreienbrock, Amely Campe.

Validation: Clara Schoneberg.

Visualization: Clara Schoneberg.

Writing – original draft: Clara Schoneberg.

Writing – review & editing: Clara Schoneberg, Jens Böttcher, Lothar Kreienbrock, Amely Campe.

References

1. Brühl-Schneider L, Hiller E, Frey T, Akimkin V, Hoferer M, Sting R. Comparative ORF and whole genome sequencing analysis of the porcine preproductive and respiratory syndrome virus (PRRSV) in routine samples reveal a recombinant virus strain. *Berliner und Münchener Tierärztliche Wochenschrift*. 2020; 133.
2. Lunney JK, Benfield DA, Rowland RR. Porcine reproductive and respiratory syndrome virus: an update on an emerging and re-emerging viral disease of swine. Elsevier; 2010.
3. Wensvoort G, Terpstra C, Pol J, Ter Laak E, Bloemraad M, De Kluyver E, et al. Mystery swine disease in The Netherlands: the isolation of Lelystad virus. *Veterinary Quarterly*. 1991; 13(3):121–30. <https://doi.org/10.1080/01652176.1991.9694296> PMID: 1835211
4. Mengeling WL, Lager KM, Vorwald AC. Diagnosis of porcine reproductive and respiratory syndrome. *Journal of Veterinary Diagnostic Investigation*. 1995; 7(1):3–16. <https://doi.org/10.1177/104063879500700102> PMID: 7779961
5. Yoon K-J, Zimmerman JJ, Swenson SL, McGinley MJ, Eernisse KA, Brevik A, et al. Characterization of the humoral immune response to porcine reproductive and respiratory syndrome (PRRS) virus infection. *Journal of Veterinary Diagnostic Investigation*. 1995; 7(3):305–12. <https://doi.org/10.1177/104063879500700302> PMID: 7578443
6. Meier WA, Galeota J, Osorio FA, Husmann RJ, Schnitzlein WM, Zuckermann FA. Gradual development of the interferon- γ response of swine to porcine reproductive and respiratory syndrome virus infection or vaccination. *Virology*. 2003; 309(1):18–31. [https://doi.org/10.1016/s0042-6822\(03\)00009-6](https://doi.org/10.1016/s0042-6822(03)00009-6) PMID: 12726723
7. Sattler T, Wodak E, Revilla-Fernández S, Schmoll F. Comparison of different commercial ELISAs for detection of antibodies against porcine respiratory and reproductive syndrome virus in serum. *BMC veterinary research*. 2014; 10(1):300. <https://doi.org/10.1186/s12917-014-0300-x> PMID: 25518885
8. Dohoo IR, Martin W, Stryhn H. Veterinary epidemiologic research. 2nd ed: AVC Incorporated Charlottetown, Prince Edward Island, Canada; 2009.
9. Valenstein PN. Evaluating diagnostic tests with imperfect standards. *American Journal of Clinical Pathology*. 1990; 93(2):252–8. <https://doi.org/10.1093/ajcp/93.2.252> PMID: 2405632
10. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *Journal of clinical epidemiology*. 1988; 41(9):923–37. [https://doi.org/10.1016/0895-4356\(88\)90110-2](https://doi.org/10.1016/0895-4356(88)90110-2) PMID: 3054000
11. Branscum A, Gardner I, Johnson W. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Preventive veterinary medicine*. 2005; 68(2–4):145–63. <https://doi.org/10.1016/j.prevetmed.2004.12.005> PMID: 15820113
12. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics*. 1980; 167–71. PMID: 7370371
13. Pouillot R, Gerbier G, Gardner IA. “TAGS”, a program for the evaluation of test accuracy in the absence of a gold standard. *Preventive veterinary medicine*. 2002; 53(1–2):67–81. [https://doi.org/10.1016/s0167-5877\(01\)00272-0](https://doi.org/10.1016/s0167-5877(01)00272-0) PMID: 11821138
14. Collins LM, Lanza ST. Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences: John Wiley & Sons; 2009.
15. Markey B, Leonard F, Archambault M, Cullinane A, Maguire D. Clinical veterinary microbiology e-book: Elsevier Health Sciences; 2013.

16. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics*. 2001; 57(1):158–67. <https://doi.org/10.1111/j.0006-341x.2001.00158.x> PMID: 11252592
17. Georgiadis MP, Johnson WO, Gardner IA, Singh R. Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2003; 52(1):63–76.
18. Black MA, Craig BA. Estimating disease prevalence in the absence of a gold standard. *Statistics in Medicine*. 2002; 21(18):2653–69. <https://doi.org/10.1002/sim.1178> PMID: 12228883
19. Yang I, Becker MP. Latent variable modeling of diagnostic accuracy. *Biometrics*. 1997;948–58. PMID: 9290225
20. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*. 1996;797–810. PMID: 8805757
21. Albert PS, McShane LM, Shih JH, Network UNCIBTM. Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. *Biometrics*. 2001; 57(2):610–9. <https://doi.org/10.1111/j.0006-341x.2001.00610.x> PMID: 11414591
22. Zhang B, Chen Z, Albert PS. Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts. *Biometrics*. 2012; 68(4):1294–302. <https://doi.org/10.1111/j.1541-0420.2012.01789.x> PMID: 23006010
23. Liu W, Zhang B, Zhang Z, Chen B, Zhou X-H. A pseudo-likelihood approach for estimating diagnostic accuracy of multiple binary medical tests. *Computational Statistics & Data Analysis*. 2015; 84:85–98.
24. Hanson TE, Johnson WO, Gardner IA. Log-linear and logistic modeling of dependence among diagnostic tests. *Preventive Veterinary Medicine*. 2000; 45(1–2):123–37. [https://doi.org/10.1016/s0167-5877\(00\)00120-3](https://doi.org/10.1016/s0167-5877(00)00120-3) PMID: 10802337
25. Schoneberg C, Kreienbrock L, Campe A. An iterative, frequentist approach for latent class analysis to evaluate conditionally dependent diagnostic tests. *Frontiers in Veterinary Science*. 2021; 8:31. <https://doi.org/10.3389/fvets.2021.588176> PMID: 33681320
26. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
27. Díaz I, Venteo Á, Rebollo B, Martín-Valls GE, Simon-Grifé M, Sanz A, et al. Comparison of two commercial enzyme-linked immunosorbent assays for the diagnosis of Porcine reproductive and respiratory syndrome virus infection. *Journal of Veterinary Diagnostic Investigation*. 2012; 24(2):344–8. <https://doi.org/10.1177/1040638711435804> PMID: 22379050
28. Gerber PF, Giménez-Lirola LG, Halbur PG, Zhou L, Meng X-J, Opiressnig T. Comparison of commercial enzyme-linked immunosorbent assays and fluorescent microbead immunoassays for detection of antibodies against porcine reproductive and respiratory syndrome virus in boars. *Journal of virological methods*. 2014; 197:63–6. <https://doi.org/10.1016/j.jviromet.2013.12.001> PMID: 24361873
29. Sipos W, Lang C, Minani A, Fischer L, Ritzmann M, Schmoll F. Comparison of two commercial ELISA systems for the detection of PRRSV-specific antibodies with a gold standard ELISA. *Wien Tierärztl Mschr-Vet Med Austria*. 2009; 96:28–33.
30. Sattler T, Pikalo J, Wodak E, Schmoll F. Ability of ELISAs to detect antibodies against porcine respiratory and reproductive syndrome virus in serum of pigs after inactivated vaccination and subsequent challenge. *BMC veterinary research*. 2016; 12(1):259. <https://doi.org/10.1186/s12917-016-0888-0> PMID: 27871292
31. Sattler T, Pikalo J, Wodak E, Schmoll F. Performance of ELISAs for detection of antibodies against porcine respiratory and reproductive syndrome virus in serum of pigs after PRRSV type 2 live vaccination and challenge. *Porcine Health Management*. 2015; 1(1):19. <https://doi.org/10.1186/s40813-015-0015-9> PMID: 28405425
32. Karniychuk U, Nauwynck H. Performance of assays for testing antibodies against porcine reproductive and respiratory syndrome virus in sera collected from swine farms in a region with an extreme virus heterogeneity. *Acta Veterinaria*. 2014; 64(1):44–51.
33. Biernacka K, Podgórska K, Tyszka A, Stadejek T. Comparison of six commercial ELISAs for the detection of antibodies against porcine reproductive and respiratory syndrome virus (PRRSV) in field serum samples. *Research in veterinary science*. 2018; 121:40–5. <https://doi.org/10.1016/j.rvsc.2018.10.005> PMID: 30316015
34. Henao-Diaz A, Giménez-Lirola L, Magtoto R, Ji J, Zimmerman J. Evaluation of three commercial porcine reproductive and respiratory syndrome virus (PRRSV) oral fluid antibody ELISAs using samples of known status. *Research in veterinary science*. 2019; 125:113–8. <https://doi.org/10.1016/j.rvsc.2019.05.019> PMID: 31181347

35. Gardner IA, Colling A, Greiner M. Design, statistical analysis and reporting standards for test accuracy studies for infectious diseases in animals: progress, challenges and recommendations. Preventive veterinary medicine. 2019; 162:46–55.
36. O'Gorman MR, Donnenberg AD. Handbook of human immunology: CRC press; 2008.
37. Gardner IA, Stryhn H, Lind P, Collins MT. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. Preventive veterinary medicine. 2000; 45(1–2):107–22. [https://doi.org/10.1016/s0167-5877\(00\)00119-7](https://doi.org/10.1016/s0167-5877(00)00119-7) PMID: 10802336

3 Übergreifende Diskussion

In dieser Arbeit wurde die Methode der LCA weiterentwickelt, um sie auch für bedingt abhängige diagnostische Tests anwenden zu können. Das klassische latente Klassenmodell wird darin um einen Term für die bedingten Abhängigkeiten der Tests erweitert. In einem Algorithmus werden dann zunächst geeignete Startwerte für alle Parameter festgelegt. Anschließend werden die bedingten Abhängigkeiten als konstant angesehen und die Testgüten sowie die Prävalenz analog zur klassischen LCA in einem Erwartungs-Maximierungs(EM)-Algorithmus (*expectation-maximization algorithm*) bestimmt. Mit diesen Werten können dann die bedingten Abhängigkeiten neu berechnet werden. Diese zwei Schritte werden abwechselnd wiederholt bis Konvergenz eintritt. In Publikation I konnte mittels verschiedener Simulationsszenarien die generelle Anwendbarkeit der Methode gezeigt werden. Anschließend wurde sie in Publikation II auf einen Datensatz bestehend aus den Ergebnissen von sechs ELISA Tests zum Nachweis von Antikörpern gegen das PRRS-Virus angewendet, um die praktische Verwendbarkeit für reale Datensätze und das dafür nötige Vorgehen zu demonstrieren.

3.1 Herausforderungen in der Anwendung der Methode

Trotz der in den Publikationen bestätigten grundsätzlichen Anwendbarkeit der vorgestellten Methode gibt es einige Aspekte, die bei ihrem Einsatz in der Praxis berücksichtigt werden müssen, um eine exakte Schätzung der Parameter zu ermöglichen. In diesem Kapitel werden die verschiedenen Herausforderungen für die praktische Anwendung näher beschrieben und erläutert, wie sie zu bewältigen sind.

3.1.1 Stichprobengröße

Zunächst ist eine ausreichend große Stichprobe notwendig, um eine möglichst gute Schätzung der Parameter sicherzustellen. Wird die Stichprobe zu klein gewählt, kann es je nach Datensatz zu einer höheren Streuung und damit zu deutlichen Abweichungen von den tatsächlichen Werten sowie zu fehlender Konvergenz des Algorithmus kommen. Die benötigte Anzahl von Beobachtungen hängt von der Anzahl der im Modell verwendeten Tests ab. Je mehr Tests es gibt, desto mehr mögliche Antwortmuster gibt es. So gibt es bei drei Tests mit zwei Antwortkategorien $2^3 = 8$ Antwortmuster oder bei sechs Tests bereits $2^6 = 64$.

mögliche Antwortkombinationen, deren Wahrscheinlichkeit im EM-Algorithmus abzuschätzen sind, so dass der erforderliche Stichprobenumfang exponentiell wachsen muss.

Es gibt bisher wenige Informationen zur statistischen Genauigkeit beziehungsweise Power sowie den darauf aufbauenden erforderlichen Stichprobenumfang für methodische Anwendungen der LCA. Für die vorgestellte Methode wurde keine solche Analyse durchgeführt und auch für andere Ansätze der LCA gibt es keine exakten Angaben zum benötigten Stichprobenumfang.

Allerdings zeigen empirischen Untersuchungen anderer zu der latenten Transitionsanalyse, einer longitudinalen Erweiterung der LCA, und zu gemischten Strukturgleichungsmodellen (*structural equation mixture models*), dass Stichprobenumfänge von etwa 300-1 000 Beobachtungen erforderlich erscheinen, um stabile Untersuchungsergebnisse zu erhalten (Collins und Wugalter 1992; Tueller et al. 2011). Eine andere Studie kam zu dem Ergebnis, dass 500 Beobachtungen eine angemessene Stichprobengröße darstellen (Finch und Bronk 2011). Dem gegenüber stehen die Ergebnisse verschiedener Monte-Carlo-Simulationen, die zu dem Schluss führten, dass bei einer größeren Anzahl von Indikatorvariablen, die einen direkten Einfluss auf die latenten Klassen haben, die Stichprobengröße von 500 unterschritten werden darf. Jedoch sollten in jedem Fall mehr als 70 Beobachtungen zur Verfügung stehen (Wurpts und Geiser 2014). Eine weitere Studie schlug einen minimalen Stichprobenumfang von 300 Beobachtungen vor, wobei dieser von weniger komplexen Modellen unterschritten werden darf (Nylund-Gibson und Choi 2018).

Diese Angaben stellen oft nur eine grobe Abschätzung dar und weichen deutlich voneinander ab, da die LCA sehr flexibel in der Zahl der latenten Klassen sowie der Form der Indikatorvariablen und damit der Komplexität des Modells ist. Die Power hängt von der Struktur des Datensatzes sowie des verwendeten Modells ab und ist abhängig vom jeweiligen Anwendungsfall. Ihre Verteilung wird von einer großen Zahl von Parametern beeinflusst, deren Zusammensetzung sich je nach Anwendung ändert und deren Werte vor der Analyse abgeschätzt werden müssen. Damit ist eine theoretische Berechnung der Power sehr aufwendig und würde gegebenenfalls in einer weiteren wissenschaftlichen Untersuchung resultieren. Eine exakte Analyse zur Genauigkeit kann in Zukunft jedoch hilfreich sein, um für Studien mit dem primären Ziel der diagnostischen Testevaluation den notwendigen Stichprobenumfang für die LCA festzulegen.

Die benötigte Stichprobengröße in dem hier vorstellten Modell dürfte aufgrund der festgelegten Zahl von latenten Klassen und Antwortkategorien der Indikatorvariablen sowie der zusätzlich berücksichtigten bedingten Abhängigkeiten von der klassischen LCA abweichen. Durch die Berücksichtigung der bedingten Abhängigkeiten besitzt das vorgestellte Modell eine höhere Komplexität, weshalb die in der Literatur empfohlenen Stichprobenumfänge auch in diesem Fall nicht unterschritten werden sollten.

In Publikation I wurden simulierte Daten verwendet. Für jedes Szenario wurden 10 000 Beobachtungen simuliert, was eine ausreichende Power sicherstellen sollte. In Publikation II standen dagegen weniger Informationen zur Verfügung. Da die Daten erst nach Abschluss der Untersuchung bereitgestellt wurden und die Durchführung der LCA nicht das primäre Ziel war, gab es keine Möglichkeit, den Stichprobenumfang zu beeinflussen. Jedoch standen mit 812 untersuchten Tieren mehr Beobachtungen zur Verfügung als in der Literatur gefordert, so dass auch hier eine ausreichende Aussagekraft gewährleistet scheint.

3.1.2 Startwerte

Darüber hinaus zeigen die durchgeföhrten Untersuchungen, dass die Güte der Schätzung bei bedingt abhängigen latenten Klassenmodellen stark von den gewählten Startwerten abhängt. Somit ist insofern Vorwissen notwendig, dass die Startwerte in realistischen Szenarien unterstellt werden sollten.

In Publikation II wurden verschiedene Wege gezeigt, diese benötigten Informationen zu erlangen: vorherige Publikationen, die die Tests evaluieren (vgl. Tabelle S5 in Anhang 2), Herstellerangaben und praktische Erfahrung der Anwender im Labor. Die Qualität der Informationsquellen hängt jedoch davon ab, wie etabliert die Tests bereits sind. Wird ein neuer Test verwendet, stehen lediglich die Ergebnisse von den Evaluationsstudien des Herstellers als Werte für die Testgüten zur Verfügung. Diese Angaben können deutlich von den tatsächlichen Werten abweichen, da die Hersteller nicht alle praktischen Anwendungsszenarien bei der Testentwicklung abbilden können. Dies konnte auch in der Studie gezeigt werden (vgl. Tabellen S3 und S4 in Anhang 2).

Für Vorinformationen zu den bedingten Abhängigkeiten muss ebenfalls entweder auf vorherige Publikationen zurückgegriffen werden, die die betrachteten Tests evaluieren und bedingte Abhängigkeiten zwischen diesen annehmen oder es muss das verwendete diagnostische Testprinzip in Kombination mit der Erfahrung aus der Praxis evaluiert werden.

Oft werden in Publikationen zur Testevaluation jedoch keine Abhängigkeiten zwischen den Tests berücksichtigt, wie im Fall von PRRS (Sipos et al. 2009; Sattler et al. 2014; Karniychuk und Nauwynck 2014; Sattler et al. 2015, 2016; Biernacka et al. 2018; Henao-Diaz et al. 2019), oder es gibt noch keine Publikationen zu neuen Tests. Gibt es also nur sehr wenige Vorinformationen, können die Startwerte stattdessen unter Berücksichtigung des verwendeten biologischen Testprinzips in Rücksprache mit den Anwendern aus dem Labor ausgewählt werden.

Um die resultierenden Werte zu bestätigen und die Stabilität des Verfahrens zu prüfen, können so mehrere realistische Startwertkombinationen ausgewählt und der Algorithmus angewendet werden. Konvergieren alle Startwerte zu denselben Parameterwerten, ist dies ein deutlicher Hinweis, dass dies die tatsächlich zugrundeliegenden Werte der Tests und der Population sind. Abweichende Werte sprechen für eine geringe Stabilität des Verfahrens für den untersuchten Datensatz: Die resultierenden Werte müssen inhaltlich in Bezug auf die Population und Tests interpretiert werden.

Durch die starke Abhängigkeit von den Vorinformationen ist es sowohl für die vorgestellte Methode als auch für alternative Ansätze der LCA wichtig, dass ausreichend Informationen zu den verwendeten Tests verfügbar sind. Eine wichtige Informationsquelle sind die Ergebnisse der Evaluationsstudien der Testhersteller oder der Labore. Diese sind jedoch oft nur schwer oder gar nicht zugänglich, was hier die Festlegung der Startwerte erschwert. Ein einfacherer Zugang zu diesen Informationen kann sowohl die Leistung der vorgestellten Methode verbessern als auch Informationen zur Qualität der Tests in verschiedenen Anwendungssituationen liefern. So wird die Möglichkeit geschaffen, geeignete diagnostische Tests für verschiedene Anwendungssituationen zu finden.

3.1.3 Anzahl der verwendeten Tests

Eine weitere notwendige Voraussetzung der Methode besteht darin, dass die Ergebnisse von mindestens drei Tests vorliegen müssen. Während zwei Tests zu wenig Informationen liefern, um die diagnostischen Testgüten und die Prävalenz im EM-Algorithmus eindeutig bestimmen zu können, kann mit mindestens drei Tests eine positive Anzahl an Freiheitsgraden für die Anwendung des Algorithmus erreicht und die Parameter so eindeutig bestimmt werden (Collins und Lanza 2009). Dies ist eine grundsätzliche Voraussetzung von allen latenten Klassenmodellen und muss bei der Studienplanung beachtet werden.

Jedoch kann es trotz Berücksichtigung dieser Voraussetzung zu Situationen kommen, in denen die entwickelte Methode nicht anwendbar ist. Beispielsweise wurde eine Studie zum Nachweis von *Coxiella (C.) burnetii* in Schafherden in Deutschland durchgeführt, um die Prävalenz des Erregers in den Herden und die Güte der verwendeten Tests zu bestimmen. In der Studie wurden insgesamt 3 367 Schafe und Ziegen aus 71 Betrieben in Deutschland mit drei diagnostischen Tests auf *C. burnetii* untersucht (Wolf et al. 2020), wodurch alle notwendigen Voraussetzungen für die Anwendung der vorstellten LCA erfüllt waren. Im praktischen Anwendungsfall führte einer der Tests zu ausschließlich negativen Ergebnissen für die gesamte Stichprobe. Offensichtlich gab es einen Fehler bei diesem neu entwickelten Test, so dass er keine zusätzlichen Informationen zum latenten Status der Tiere lieferte. Damit standen de facto nur noch die Ergebnisse von zwei diagnostischen Tests zur Verfügung. In diesem Fall ist der EM-Algorithmus aufgrund der negativen Zahl von Freiheitsgraden nicht in der Lage, eindeutige Ergebnisse zu bestimmen, wodurch die Methode instabil wird.

Die vorgestellte Methode besitzt jedoch grundsätzlich keine Einschränkung in der maximal zu verwendenden Zahl von Tests. Deshalb wurde sie in Publikation II auf die Ergebnisse von sechs ELISAs zum Nachweis von Antikörpern gegen PRRS angewendet, um die Anwendbarkeit der Methode auf reale Daten zu prüfen. Häufig wird zwar der PRRS X3 ELISA des Herstellers Idexx als Goldstandardtest für den Nachweis von Antikörpern gegen das zugrundeliegende Virus angegeben (Chen et al. 2013; Cong et al. 2013), jedoch wurde in verschiedenen Studien bereits eine suboptimale Testgüte dieses Tests bestimmt (Karniychuk und Nauwynck 2014; Sattler et al. 2016; Biernacka et al. 2018). Damit ist dies der ideale Anwendungsfall der Methode, da mit sechs ELISAs eine ausreichende Anzahl von Tests vorhanden ist, wegen des gleichen Testprinzips eine hohe Abhängigkeit vorliegt und es keinen optimalen Goldstandardtest für den Nachweis von Antikörpern gegen PRRS gibt. Die entwickelte Methode war in der Lage, alle Testgüten und die Prävalenz für die untersuchte Stichprobe zu bestimmen.

Somit benötigt die vorgestellte Methode die gleiche minimale Menge von Daten wie die klassische LCA und ist für die Anwendung auch auf eine größere Zahl diagnostischer Tests geeignet.

3.1.4 Stabilität der Schätzung

Die Stabilität des Verfahrens und die Güte der Schätzung hängen von verschiedenen Faktoren ab. Die Anzahl der verwendeten Tests beeinflusst die Komplexität des Modells. Grundsätzlich gibt es zwar, wie in Kapitel 3.1.3 beschrieben, keine Einschränkungen in der maximalen Zahl von Tests im Modell. Jedoch gilt auch, dass je mehr Tests verwendet werden, desto mehr Daten sind vorhanden und desto mehr Parameter müssen geschätzt werden. Formel (a) gibt an, wie viele Parameter in Abhängigkeit der Anzahl der verwendeten Tests n im Modell geschätzt werden.

$$P = 1 + 2n + 2 \sum_{i=2}^n \binom{n}{i} \quad (\text{a})$$

Dagegen gibt es 2^n mögliche Antwortkombinationen, die beobachtet werden können. Damit wächst die Zahl der zu schätzenden Parameter mit steigender Anzahl verwendeter Tests stärker als die Zahl der verwendeten Daten. Auch kann es deutlich schwieriger sein, die Startwerte für die Abhängigkeiten mehrerer Tests zu bestimmen. Während paarweise Abhängigkeiten noch sehr praxisnah zu verstehen sind, stellen Abhängigkeiten höherer Ordnung eine größere Herausforderung dar. Somit kann es bei einer größeren Anzahl von verwendeten Tests zu höheren Unsicherheiten kommen.

Auch das Verhältnis der positiven zu den negativen Beobachtungen kann die Güte der Schätzung beeinflussen, wie die Simulationsstudien in Publikation I zeigen. Bei einem sehr unausgeglichenen Verhältnis der latenten Klassen zueinander (3% Prävalenz) gab es in der Simulation schon bei leicht abweichenden Startwerten große Abweichungen der Ergebnisse (vgl. Tabellen S3 und S8 in Anhang 1). Ein ähnliches Verhalten wurde auch bei anderen Modellen der LCA beobachtet (McAloon et al. 2019). Der Grund dafür ist der stärkere Einfluss der Abweichungen bei geringer Prävalenz. Wird die Prävalenz zum Beispiel um 1% überschätzt, hat dies bei einem ausgeglichenen Verhältnis der latenten Klassen zueinander kaum Auswirkungen. Bei einer Prävalenz von 3% entspricht dies aber bereits einer Abweichung von einem Drittel der Beobachtungen. So führt ein unausgeglichenes Verhältnis der latenten Klassen besonders in der kleineren Klasse zu starken Abweichungen der Parameterwerte, während es kaum einen Effekt in der anderen latenten Klasse gibt. Dies stellt ein Problem dar, da viele Infektionskrankheiten nur mit einer geringen Prävalenz einhergehen (zum Beispiel (Coulibaly und Yameogo 2000; Berger-Schoch et al. 2011; Nielsen und Toft 2009)).

In Publikation II wurde zunächst eine Population von Schweinen mit unterschiedlichen Immunstatus untersucht. Dabei führten alle Startwerte, die zum Teil sehr unterschiedlich waren, zum gleichen Ergebnis (Tabelle S3 in Anhang 2). Bei der Analyse einer Untergruppe dieser Daten, welche ausschließlich geimpfte Tiere enthielt, ergaben sich stark abweichende Ergebnisse bei verschiedenen Startwerten (Tabelle S4 in Anhang 2). Ein Grund für die geringere Stabilität des Verfahrens kann der erhöhte Anteil positiver Tiere in der geimpften Gruppe und damit ein stärkeres Ungleichgewicht der zwei latenten Klassen sein. Eine andere mögliche Ursache kann die größere Homogenität der Untergruppe im Vergleich zum gesamten Datensatz darstellen. Alle Tiere in der untersuchten Untergruppe waren gegen den europäischen Genotyp 1 des Virus geimpft, während im gesamten Datensatz Tiere enthalten waren, die gegen den europäischen Genotyp, den amerikanischen Genotyp 2 oder gegen beide Genotypen geimpft waren. Ebenso waren Tiere enthalten, die keine Impfung erhalten hatten. Somit gab es in der Untergruppe homogener Immunstatus der Tiere, wodurch einige Antwortmuster anteilig häufiger beobachtet wurden, andere Antwortkombinationen, die im gesamten Datensatz vorhanden waren, hingegen gar nicht mehr. Dies kann die Konvergenz des Algorithmus zu den richtigen Parameterwerten erschweren, da weniger Daten zur Verfügung stehen. Mithilfe zusätzlicher Simulationsdatensätze kann dieses Verhalten ausführlicher untersucht werden. Beispielsweise ist noch nicht bekannt, ab welchem Anteil übereinstimmend geimpfter Tiere diese Schwierigkeiten auftreten können. Bei der Anwendung des Modells auf einen Datensatz mit teilweise oder komplett geimpften Tieren, sollte dies jedoch berücksichtigt werden.

Es gibt also viele Faktoren, die das Verfahren beeinflussen, zum Beispiel die verwendeten Startwerte, die Anzahl der betrachteten Tests oder der Immunstatus der untersuchten Population. Einige Punkte davon sind bereits aus anderen Ansätzen bekannt, welche die bedingten Abhängigkeiten zwischen den Tests berücksichtigen, andere sind noch nicht untersucht worden und benötigen tiefergehende Analysen. Daher ist es wichtig, vor der Auswahl der Tests und der Durchführung der Analyse diese Punkte zu berücksichtigen, um mögliche Schwierigkeiten erkennen zu können.

3.1.5 Parameterschranken

Die in dieser Arbeit vorgestellte Methode ist, wie in Kapitel 3.1.2 diskutiert, von den gewählten Startwerten abhängig. Falsche Annahmen zu den diagnostischen Testgüten und den Abhängigkeiten können dazu führen, dass der Algorithmus stark von den wahren

Parametern abweichende Werte ermittelt. Dies kann in den Situationen, in denen wenig Informationen zu den verwendeten Tests vorhanden sind, ein Problem darstellen und die praktische Nutzbarkeit der Methode erheblich einschränken.

Eine Lösung für diese Problematik wird in Publikation I in Form von Parametergrenzen vorgestellt. Überschreitet ein Parameter die für ihn festgelegte Schranke, wird er im nächsten Iterationsschritt automatisch korrigiert. Zusätzlich zu den exakt gewählten Startwerten können so Intervalle für die Parameter festgelegt werden, die dem Anwender realistisch erscheinen. Zum Beispiel können die Abhängigkeiten der Tests auf positive Werte eingegrenzt oder eine untere Abschätzung der diagnostischen Testgüten getroffen werden. Die Simulationsstudien zeigen, dass diese Grenzen eine Konvergenz zu stark von den wahren Parametern abweichenden Werten verhindern können.

Die Hinzunahme dieser Grenzen hat den Vorteil, dass keine exakten Angaben zu den Werten gemacht werden müssen, die Konvergenz zu stark abweichenden Werten aber dennoch verhindert wird. Jedoch wird auch dafür Vorwissen zu den Tests und der untersuchten Population benötigt. Ohne jegliches Vorwissen sind die Grenzen nicht anwendbar. In einem solchen Fall können auch keine angemessenen Startwerte gewählt werden, was die Anwendung einer bedingt abhängigen LCA erschwert.

Ein weiteres Problem in diesem Zusammenhang können falsche Vorinformationen darstellen. Werden andere Studien mit abweichenden Studienpopulationen und anderen statistischen Auswertungsmethoden für die Abschätzung der Startwerte verwendet, können die resultierenden Werte stark von den wahren Parametern in der untersuchten Stichprobe abweichen (Gardner et al. 2019). Werden diese Informationen als Grundlage für die Parameterschranken verwendet, ist es möglich, dass die tatsächlichen Parameterwerte ausgeschlossen werden. Dann kann der Algorithmus diese Werte nicht mehr erreichen und konvergiert zu falschen Parametern oder gegebenenfalls gar nicht. Dieses Risiko kann durch ein sehr breit gewähltes Intervall möglicher Werte reduziert werden.

Andererseits gilt, je breiter das Intervall möglicher Werte gewählt wird, desto mehr Werte werden eingeschlossen und desto geringer ist der Effekt dieser Grenzen. So zeigen die Simulationsstudien in Publikation I, dass sehr breit gewählte Parametergrenzen nur bei sehr stark von den tatsächlichen Werten abweichenden Ergebnissen einen Effekt haben. Deshalb muss bei dieser Art der Parameterbeschränkung ein Kompromiss zwischen zu starken Einschränkungen einerseits und zu geringen Einschränkungen andererseits gefunden werden.

Alternativ kann auch eine andere Art der Parameterbeschränkung gewählt werden. Statt einer starr gewählten Grenze, welche die Gefahr des Ausschlusses der echten Werte beinhaltet, kann eine Straffunktion dem Modell hinzugefügt werden. Die Parametergrenzen werden dabei als Randbedingungen festgelegt und mit einem Strafparameter gewichtet.

Im vorgestellten Modell wird angenommen, dass es eine latente Variable mit zwei Klassen gibt, die durch M beobachtete Variablen gemessen wird. Die Beobachtungen $Y_i = (Y_{i1}, \dots, Y_{iM})$, $i = 1, \dots, N$ repräsentieren das individuelle Antwortmuster für die Beobachtung Y_{im} mit den möglichen Werten $r_m = 0, 1$. Die Wahrscheinlichkeit der Zugehörigkeit zur latenten Klasse c kann ausgedrückt werden als γ_c mit $\sum_{c=0}^1 \gamma_c = 1$, und die Wahrscheinlichkeit der Antwort r_m auf die Variable m in Klasse c kann formuliert werden als $\rho_{m,r_m|c}$. Sei $I(\cdot)$ die Indikatorfunktion. Dann kann die Likelihood-Funktion der Parameter γ und ρ für die Beobachtungen Y in der Form (b) geschrieben werden.

$$L(\gamma, \rho | Y) = \sum_{i=1}^N \sum_{c=0}^1 \gamma_c \left(\prod_{m=1}^M \prod_{r_m=0}^1 \rho_{m,r_m|c}^{I(Y_{im}=r_m)} + \eta_{r_1, \dots, r_M}^c \right) \quad (b)$$

Dabei ist η_{r_1, \dots, r_M}^c ein Term für den Einfluss der insgesamt $\sum_{i=2}^M \binom{M}{i}$ bedingten Abhängigkeiten aller M Tests in Klasse c auf die Likelihood-Funktion des jeweiligen beobachteten Antwortmusters. Die bedingten Abhängigkeiten werden als bedingte Kovarianzen zwischen den Testergebnissen in den zwei latenten Klassen formuliert. Sie werden jeweils als Differenz zwischen dem beobachteten Anteil übereinstimmender richtiger Antworten der Tests in latenter Klasse c und dem erwarteten entsprechenden Anteil in Klasse c unter der Annahme der bedingten Unabhängigkeit berechnet.

Die Straffunktion $P(\gamma, \rho)$ kann dann abhängig von den gewünschten Einschränkungen für einen oder mehrere Parameter innerhalb einer latenten Klasse festgelegt werden, so dass die neue Likelihood-Funktion die Form (c) besitzt.

$$L(\gamma, \rho | Y) = \sum_{i=1}^N \sum_{c=0}^1 \gamma_c \left(\prod_{m=1}^M \prod_{r_m=0}^1 \rho_{m,r_m|c}^{I(Y_{im}=r_m)} + \eta_{r_1, \dots, r_M}^c - P(\gamma, \rho) \right) \quad (c)$$

Wird zum Beispiel angenommen, dass die Wahrscheinlichkeit einer positiven Antwort des Tests 1 in der positiven latenten Klasse, also die Sensitivität von Test 1, kleiner als 90% ist, dann hat die Straffunktion P die Form $P(\rho_{1,1|1}) = \max(0, \rho_{1,1|1} - 0,9)$. Werte größer 90%

führen damit zu einem positiven Wert der Straffunktion, so dass der Wert der Likelihood-Funktion reduziert wird und das Maximum der Funktion sich ändert.

Soll die Bestrafung des Überschreitens der festgelegten Grenzen stärker ausfallen, dann kann die Funktion zum Beispiel mit $P(\rho_{1,1|1}) = \max(0, \rho_{1,1|1} - 0,9)^2$ zu einer quadratischen Verlustfunktion umgewandelt werden.

Diese Art der Parameterbeschränkung hat gegenüber festen Grenzen den Vorteil, dass Werte außerhalb des festgelegten Bereichs nicht grundsätzlich als mögliche Ergebnisse ausgeschlossen werden. Liegt der tatsächliche Parameter nur geringfügig außerhalb des definierten Wertintervalls, ist der Algorithmus mit den festen Grenzen nicht mehr in der Lage, zu diesem Wert zu konvergieren. Wird stattdessen die Straffunktion angewendet, deuten die Beobachtungen weiter auf den tatsächlichen Parameter hin und der Algorithmus ist in der Lage, zu diesem Wert zu konvergieren. Ein weiterer Vorteil der Straffunktion ist die Berücksichtigung der Stärke der Abweichung vom zu Beginn definierten Wertebereich. Leichte Abweichungen haben nur einen geringen Einfluss auf die Likelihood-Funktion, während stärkere Abweichungen einen größeren Einfluss haben. So wird die Wahrscheinlichkeit mehrerer lokaler Maxima der Likelihood-Funktion und damit nicht eindeutiger Lösungen, wie in Publikation I beschrieben, reduziert.

Diese Methode der Parametereinschränkung hat jedoch auch einige Schwächen. Zunächst müssen Wertebereiche für die Parameter mit hoher Sicherheit bekannt sein. Abhängig von der Art des Vorwissens und der Menge an Parametern, zu denen es Vorwissen gibt, werden diese der Likelihood-Funktion hinzugefügt. Dies wiederum erhöht die Komplexität des Modells und die Menge der Informationen, die der Anwender zu Beginn der Berechnung bereitzustellen hat, was der einfachen Anwendbarkeit als definiertem Ziel der Methode entgegensteht.

Trotzdem kann die Parameterbeschränkung in einzelnen Fällen bei sehr speziellen Vorinformationen hilfreich sein. Jedoch sollte sie nur als zusätzliche Möglichkeit für eine präzisere Schätzung betrachtet werden und der Schwerpunkt sollte immer auf den Startwerten liegen. Denn gut gewählte Startwerte führen zu einer Konvergenz gegen die tatsächlichen Parameterwerte ohne Parameterschranken, wie die Publikationen I und II gezeigt haben.

3.2 Anwendungsbereich der Methode

Werden die im letzten Kapitel beschriebenen Aspekte bei der Anwendung der neu entwickelten Methode der LCA berücksichtigt, kann sie in einem weiten Gebiet praktisch angewendet werden. In diesem Kapitel wird der Anwendungsbereich der Methode näher beschrieben, auf mögliche Einschränkungen der Anwendbarkeit eingegangen und die Praktikabilität des entwickelten Algorithmus bewertet.

3.2.1 Anwendungsbereich in der medizinischen diagnostischen Testevaluation

Die vorgestellte iterative, frequentistische LCA wurde entwickelt, um bedingt abhängige diagnostische Tests ohne Goldstandard zu evaluieren. Durch den Verzicht auf einen Goldstandard hat die Methode einen sehr großen Anwendungsbereich. Während viele Goldstandardmethoden zeitintensiv oder invasiv unter Laborbedingungen durchgeführt werden (zum Beispiel mehrwöchige kulturelle Anzucht des Erregers oder Autopsie des Wirtes), kann die Fähigkeit der Tests mit der vorgestellten Methode unter realen Bedingungen und mit einer größeren Stichprobe im Feld untersucht werden. Einflüsse der Population werden so direkt abgebildet und es wird eine realistischere Einschätzung der praktischen Anwendbarkeit der Tests getroffen.

In dem vorgestellten Modell wird angenommen, dass jeder untersuchte diagnostische Test prinzipiell eine binäre Einteilung der Proben vornimmt und somit jeder Probe stets ein negatives oder positives Ergebnis zuzuordnen ist. Dies kann generell für jeden diagnostischen Test angenommen werden, auch wenn zunächst mehr als zwei Kategorien angenommen werden oder auch (semi-)quantitative diagnostische Testergebnisse ermittelt werden. Die Methode teilt die Beobachtungen einer Stichprobe abhängig vom Antwortmuster der verwendeten Tests in eine positive und eine negative Klasse ein und bestimmt entsprechend die Prävalenz. Anhand der abweichenden Testergebnisse werden dann die Sensitivitäten und Spezifitäten der Tests berechnet.

Die vorgestellte Methode basiert auf der in Collins und Lanza (2009) beschriebenen klassischen LCA für bedingt unabhängige diagnostische Tests, zu der ein Term für die bedingten Abhängigkeiten hinzugefügt wurde. Werden bedingt unabhängige Tests ausgewertet, können diese Abhängigkeitsterme auf null gesetzt und die Parameter dann mit diesen Werten berechnet werden. Die Ergebnisse der Simulationsstudien in Publikation I zeigen, dass die Methode auch in diesem Fall in der Lage ist, die wahren Parameterwerte zu

bestimmen. Folglich kann die Methode für alle diagnostischen Tests, unabhängig von der Stärke ihrer Abhängigkeit, verwendet werden. Dabei werden alle Diagnostikverfahren, die einem Individuum oder einer Probe einen eindeutigen und übereinstimmenden positiven oder negativen Status zuordnen, als diagnostische Tests betrachtet. Beispielsweise kann auch die diagnostische Güte von körperlichen Untersuchungen analysiert werden, ebenso wie die Fähigkeit Diagnosen anhand bildgebender Verfahren zu stellen.

3.2.2 Biologische Restriktionen

Unter das genannte Anwendungsgebiet der diagnostischen Testevaluation fallen insbesondere Tests, die auf das Vorliegen einer Infektionskrankheit prüfen, da sie eine eindeutige Zuordnung des Krankheitsstaus vornehmen. Bei der Auswahl der diagnostischen Tests für die Analyse muss aber berücksichtigt werden, welche Infektionsphase die Tests nachweisen. Werden beispielsweise Antigentests (zum Beispiel eine PCR-Analyse) gemeinsam mit Antikörpertests evaluiert, weisen diese unterschiedliche biologische Infektionsstadien nach und die Methode sollte in ihrer eigentlichen Bedeutung des identischen Untersuchungsziels nicht angewendet werden.

Während Antigentests den Erreger selbst oder Teile seiner DNA nachweisen, prüft der Antikörpertest auf die Immunantwort des Körpers in Form von vorhandenen Antikörpern. Jedoch treten diese Zustände zu unterschiedlichen Zeitpunkten einer Infektion ein. So ist der Erreger zu Beginn der Infektion nachweisbar und kann zu einem späteren Zeitpunkt kaum oder gar nicht mehr vorhanden sein. Antikörper bilden sich erst nach dem Kontakt mit dem Erreger (Markey et al. 2013). Damit messen die Tests nicht exakt dieselbe latente Infektionsphase der Tiere.

Diese abweichenden Ergebnisse der Tests basieren auf den biologischen Abläufen einer Infektionskrankheit. In dem latenten Klassen-Ansatz wird hingegen angenommen, dass es zwei eindeutige Klassen gibt, zu denen jedes Individuum durch die Tests zugeordnet werden kann. Diese Voraussetzung muss bei allen Modellen der LCA für eine unverzerrte Schätzung der Parameter erfüllt sein.

Im Fall von Antigen- und Antikörpertests gibt es hingegen vier Klassen, nach denen die Individuen unterteilt werden müssen:

1. Nur Antigene vorhanden

2. Sowohl Antigene als auch Antikörper vorhanden
3. Nur Antikörper vorhanden
4. Weder Antigene noch Antikörper vorhanden

Für eine gemeinsame LCA von Antigen- und Antikörpertests müssen also vier latente Klassen betrachtet werden, damit jedes Individuum einem der vier genannten Zustände zugeordnet werden kann. Die Antwortwahrscheinlichkeiten der Tests in den vier Klassen geben dann die Testgüten einmal mit und einmal ohne Vorliegen des anderen Infektionsnachweises an. Anhand des Anteils der Individuen in den zwei Antigen- beziehungsweise Antikörper-positiven Klassen kann die Prävalenz beider Zustände getrennt bestimmt werden.

Jedoch ist bei dieser Vergrößerung der Anzahl der latenten Klassen die gleichzeitige Vergrößerung der Anzahl der zu berechnenden Parameter zu berücksichtigen. Dadurch verdoppelt sich die Zahl der Antwortwahrscheinlichkeiten. Auch die Zahl der zu schätzenden bedingten Abhängigkeiten erhöht sich deutlich, je nach Anzahl der bedingt abhängigen Tests. Somit führt eine Vergrößerung der Zahl der latenten Klassen zu einem erhöhten Rechenaufwand, der die benötigte Stichprobe und die Anzahl der zu wählenden Startparameter vergrößert. Auch zeigen die Ergebnisse der Simulationsstudien in Publikation I, dass die Methode bei sehr ungleicher Größe der latenten Klassen zu Stabilitätsproblemen des Algorithmus führen kann. Somit führt die gleichzeitige Auswertung von diagnostischen Tests mit unterschiedlichen biologischen Testprinzipien sowohl zu einem erhöhten Aufwand als auch zu höheren Unsicherheiten und sollte möglichst vermieden werden.

3.2.3 Weitere Anwendungsbereiche

Auch außerhalb der (veterinär-)medizinischen Diagnostik lassen sich Anwendungsgebiete für das vorgestellte Verfahren finden. In anderen Gebieten der Epidemiologie kann eine Anwendung der vorgestellten LCA ebenfalls hilfreich sein, wenn anhand verschiedener abhängiger Indikatoren latente Strukturen identifiziert werden sollen. Beispielsweise können landwirtschaftliche Betriebe anhand mehrerer beobachteter Variablen in latente Klassen bezüglich ihres Tierwohlstatus eingeteilt oder Risikogruppen für die Infektion mit einer bestimmten Erkrankung definiert werden.

In der Soziologie gibt es ebenfalls Bereiche, in denen eine bedingte Abhängigkeit berücksichtigt werden kann. Wird beispielsweise eine LCA basierend auf Angaben in einem Fragebogen durchgeführt, können die Teilnehmer auf verschiedene Fragen aus den gleichen

Gründen falsche Angaben machen. Damit gibt es bestimmte Antwortmuster in einer Klasse besonders häufig. Diese Antwortmuster werden in der klassischen LCA überbewertet. Dadurch sind die daraus berechneten bedingten Antwortwahrscheinlichkeiten verzerrt, was zu einer falschen latenten Klassenzugehörigkeit der einzelnen Beobachtungen führen kann. Die Einbeziehung bedingter Abhängigkeiten kann dies verhindern und dadurch auch gegebenenfalls die Trennbarkeit der latenten Klassen verbessern.

Auch für andere Anwendungsbereiche, wie zum Beispiel die Wirtschaftswissenschaften, kann die Berücksichtigung möglicher Abhängigkeiten sinnvoll sein. Jedoch muss in diesen Fällen die Vergrößerung der Zahl der latenten Klassen sowie der Anzahl der Antwortkategorien ermöglicht und die damit verbundene höhere Komplexität des Modells berücksichtigt werden. Viele dieser Anwendungsbereiche benötigen mehr als die zwei latenten Klassen und Antwortkategorien, welche in der Diagnostik üblich sind. Dies erhöht die Zahl der zu schätzenden Parameter deutlich und vergrößert damit auch den Aufwand für die Festlegung der Startwerte. Ein Vorteil dieser Anwendungsbereiche ist jedoch, dass sie größere Datensätze bereitstellen können und dabei nicht den regulatorischen Einschränkungen der Diagnostik unterliegen. Somit ist eine exakte Schätzung einer größeren Zahl von Parametern möglich.

3.2.4 Technische Anwendbarkeit (Softwarerealisation)

Der verwendete Algorithmus wurde in R (Version 3.5.0; R Development Core Team 2019) implementiert. Die Methode besteht aus drei Hauptfunktionen: Zunächst wird ein EM-Algorithmus angewendet, der die Testgüten und die Prävalenz analog zur klassischen LCA berechnet. Die als Parameter übergebenen bedingten Abhängigkeiten werden dabei als konstante Werte betrachtet. Eine zweite Funktion verwendet die vorliegenden Werte der Testgüten, der bedingten Abhängigkeiten und der Prävalenz, um zu bestimmen, welchen Anteil jede latente Klasse an den Beobachtungen aller Antwortmuster besitzt. Anschließend wird die Differenz zu den anhand der verwendeten Testgüten erwarteten Anteilen berechnet und daraus die bedingten Abhängigkeiten neu bestimmt. Eine dritte Funktion ruft diese beiden Methoden abwechselnd auf und prüft nach jedem Schritt, ob es zu einer Konvergenz gekommen ist, also ob die Differenz der Parameter zweier aufeinanderfolgender Iterations-schritte kleiner 0,00001 ist, oder ob die maximale Anzahl von 1 000 Iterationsschritten erreicht wurde, so dass die Berechnung abgebrochen wird.

Zunächst wurde der Algorithmus für die Untersuchungen in Publikation I für drei Tests implementiert. Für die Analyse in Publikation II wurde das Programm anschließend für eine beliebige Anzahl von Tests erweitert. In dieser Version muss der Anwender zu Beginn die Funktionen einmalig initialisieren und die Zahl der verwendeten Tests sowie die zugehörigen Startwerte eingeben. Anschließend berechnet das Programm automatisch die Ergebnisse der LCA für bedingt abhängige Tests. Weitere Startwerte können durch ein wiederholtes Aufrufen der Gesamtfunktion mit geänderten Eingabeparametern direkt verwendet werden. Somit ist die Anwendung der Methode direkt möglich und es werden ausschließlich Angaben zur Anzahl der Tests und der Startwerte benötigt, um den Algorithmus anzuwenden. Für eine Änderung der Zahl der Tests müssen die neuen Startwerte übergeben und die Anzahl der zu schätzenden Tests angegeben werden. Folglich ist eine leichte Nutzbarkeit für alle Anwendungsfälle gesichert.

Jedoch wird in der jetzigen Ausführung des Programms vorausgesetzt, dass der Nutzer mit der Anwendung von Skriptsprachen wie R vertraut ist. Dies ist auch bei Implementierungen von anderen Ansätzen der LCA der Fall. Auch dort werden Kenntnisse über den Aufbau der verwendeten Skriptsprache (zum Beispiel OpenBUGS (Lunn et al. 2009), PROC LCA in SAS (Lanza et al. 2007) oder TAGS in R (Pouillot et al. 2002)) vorausgesetzt. Die bisherigen Anwender der LCA sind demnach mit dem Umgang der Programme vertraut, so dass dies zunächst kein Hindernis in der Anwendbarkeit der Methode darstellt. Um die entwickelte Methode einer größeren Gruppe von Anwendern zugänglich zu machen, sollte die Methode zukünftig mithilfe der Entwicklung einer grafischen Nutzeroberfläche (*Graphical User Interface*) oder einer Webanwendung, beispielsweise RShiny, erweitert werden.

3.3 Vergleich mit Verfahren der Bayes-Statistik

Die in dieser Arbeit adaptierte Methode der konfirmatorischen Analyse zur Evaluation von diagnostischen Test ohne Goldstandard mithilfe der LCA stellt eine Alternative zur Anwendung von Modellen der Bayes-Statistik dar. Beide Ansätze sind dabei auf Vorinformationen zu den Parametern angewiesen. Bei der latenten Klassenanalyse für bedingt abhängige Tests wird das Vorwissen in der neu entwickelten Methode in Form von Startwerten für die diagnostischen Testparameter sowie deren Abhängigkeiten benötigt. Auch wenn dies gegebenenfalls viele Informationen sind, so hat dies den Vorteil, dass jeweils ein einzelner „Erfahrungswert“ in Zusammenarbeit mit den beteiligten diagnostischen Laboren

vorgegeben werden muss. Hier geht dann also unmittelbar die Alltagserfahrung der Labore ein.

Im Gegensatz dazu fordern die Methoden der Bayes-Analyse Vorgaben in Form von A-priori-Verteilungen, welche „Struktur und Grad der Unsicherheit“ widerspiegeln sollen. Diese Modellierung der Unsicherheit durch eine A-priori-Verteilung im Bayesschen Ansatz kann als großer Vorteil angesehen werden. Die strukturellen Vorgaben einer solchen Verteilung können den praktischen Anwender jedoch besonders bei wenig oder ungenauem Vorwissen gegebenenfalls überfordern, sodass es gegebenenfalls einfacher sein kann, sich auf einen einzigen „Erfahrungswert“ für die Parameter festzulegen als einen Wertebereich anzugeben.

Sind keine Informationen zu den verwendeten Tests erhältlich, werden im Bayesschen Ansatz oft nicht informative A-priori-Verteilungen verwendet, also zum Beispiel eine Gleichverteilung auf dem Einheitsintervall oder eine Normalverteilung (Berger 1985). Diese enthalten keine Informationen zu den Parametern und entsprechen so nicht der Grundidee der Verwendung bestehenden Vorwissens. So zeigen die Simulationsstudien in Publikation I, dass das Modell im Falle einer nicht informativen A-priori-Verteilung teilweise, abhängig von der betrachteten Population, zu stark von den tatsächlichen Parametern abweichenden Werten konvergiert (vgl. Tabellen S15-S26 in Anhang 1). Beispielsweise führte die Annahme von Werten von etwa 50% für alle Parameter im ersten Simulationsszenario zwar zu einer Konvergenz des Verfahrens, die Werte für die Prävalenz wurden jedoch um etwa 50% überschätzt und die der Testgüten ähnlich stark unterschätzt (Tabelle S20 in Anhang 1).

Für den vorgestellten Ansatz hingegen werden bei geringem Vorwissen eventuell stark von den Parameterwerten abweichende Startwerte ausgewählt, die dann in den durchgeführten Simulationsstudien meist zu falschen Parameterwerten konvergierten. Um dies zu verhindern, können mehrere Startwertdatensätze anhand der verwendeten Testprinzipien und Erfahrung in der Anwendung dieser oder ähnlicher Tests bestimmt werden. Übereinstimmende beziehungsweise abweichende Ergebnisse liefern Hinweise zur Stabilität des Verfahrens für den untersuchten Datensatz und den tatsächlichen Parameterwerten. Jedoch kann es hier Schwierigkeiten geben, Startwerte für die bedingten Abhängigkeiten der Tests festzulegen. Die Anwender können Probleme damit haben, die Abhängigkeiten zwischen den Tests zu quantifizieren, da diese, anders als Testgüten oder Prävalenzen, nicht zu den bekannten Maßen der Diagnostik gehören. Dafür kann eine Anleitung zur Einordnung der Werte in einen praktischen Kontext hilfreich sein. Außerdem werden als Startwerte nur ungefähre Angaben benötigt, so dass eine grobe Einordnung durch die Experten ausreichend ist.

Die durchgeführten Simulationsstudien zeigten ebenfalls, dass beide Ansätze, welche die bedingten Abhängigkeiten berücksichtigen, bei geeignet gewählten Startwerten beziehungsweise informativen A-priori-Verteilungen zu genaueren Ergebnissen kamen als die klassische LCA. Lediglich bei unabhängigen Tests gab es keinen Unterschied zwischen den drei Methoden. Bei Szenarien mit bedingt abhängigen Indikatorvariablen hatten die Bayessche LCA und der neu entwickelte Ansatz in wenigen Fällen Schwierigkeiten mit stark von den simulierten Parametern abweichenden Ergebnissen oder fehlender Konvergenz (vgl. Tabellen S3-S26 in Anhang 1). Diese traten jeweils bei unterschiedlichen Simulations-szenarien und Annahmen zum Vorwissen auf, sodass die Ursache für diese Probleme nicht im Simulationsszenario, sondern in der verwendeten Methode zu liegen scheint. Jedoch kamen beide Methoden in den meisten Szenarien übereinstimmend zu genauen Schätzungen der Parameterwerte. Tabelle 1 stellt die Ergebnisse der Schätzung der drei in Publikation I angewendeten Methoden der LCA für die fünf untersuchten Simulationsszenarien zusammenfassend gegenüber.

Tabelle 1: Zusammengefasste Ergebnisse der drei Ansätze der LCA in den fünf in Publikation I simulierten Szenarien bei geeigneten Startwerten. Die vollständigen Resultate sind in den Tabellen S3-S26 in Anhang 1 aufgeführt.

Simulationsszenario	Klassische LCA	Bayessche LCA	Neue iterative LCA
Drei unabhängige Tests mit hohen Testgüten bei mittlerer Prävalenz	Genaue Schätzung der Parameter (max. Abweichung 1%)	Genaue Schätzung der Parameter (max. Abweichung 7%)	Genaue Schätzung der Parameter (max. Abweichung 4%)
Zwei stark abhängige Tests mit mittlerer Testgüte und ein unabhängiger Test mit hoher Testgüte bei hoher Prävalenz	Starke Abweichung der geschätzten Parameter zu den wahren Werten (max. 20%)	Genaue Schätzung der Parameter (max. Abweichung 7%)	Genaue Schätzung der Parameter (max. Abweichung 7%)
Zwei stark abhängige Tests mit mittlerer Testgüte und ein unabhängiger Test mit hoher Testgüte bei niedriger Prävalenz	Sehr starke Abweichung der geschätzten Parameter zu den wahren Werten (max. 77%)	Meist genaue Schätzung der Parameter; im Fall von über- bzw. unterschätzten Testgüten mittlere Abweichungen von max. 15%	Meist genaue Schätzung der Parameter; im Fall von unterschätzten Abhängigkeiten oder überschätzten Testgüten zum Teil starke Abweichungen von max. 86%

Drei Tests mit mittlerer Abhangigkeit und mittleren Testguten bei hoher Pravalenz	Starke Abweichung der geschatzten Parameter zu den wahren Werten (max. 13%)	Meist genaue Schatzung der Parameter; im Fall von unterschatzten Testguten starke Abweichungen von max. 20%	Genaue Schatzung der Parameter (max. Abweichung 6%)
Drei Tests aus einem realen Datensatz mit variierenden Testguten und Abhangigkeiten bei mittlerer Pravalenz	Starke Abweichung der geschatzten Parameter zu den wahren Werten (max. 20%)	Meist genaue Schatzung der Parameter (max. Abweichung 6%); fehlende Konvergenz in einem Fall	Meist genaue Schatzung der Parameter (max. Abweichung 8%)
Zwei stark abhangige Tests und ein unabhangiger Test bei niedriger Pravalenz und Parameterschranken	Sehr starke Abweichung der geschatzten Parameter zu den wahren Werten (max. 77%)	Meist genaue Schatzung der Parameter, jedoch mittlere Abweichungen bei uber- bzw. unterschatzten Testguten (max. Abweichung 15%)	Meist genaue Schatzung der Parameter; im Fall von unterschatzten Abhangigkeiten oder overschatzten Testguten zum Teil starke Abweichungen von max. 58%

Insgesamt sind beide Ansatze, sowohl das hier vorstellte Modell als auch die Bayessche Methode, geeignet zur diagnostischen Testevaluation bedingt abhangiger Tests. Sie nutzen jedoch unterschiedliche statistische Konzepte und sind daher nicht direkt vergleichbar. Je nach Ziel der Analyse und Form der vorhandenen Vorinformationen kann jeweils einer der Ansatze einfacher oder schneller anwendbar sein. Damit stellt die in dieser Arbeit vorstellte frequentistische LCA keinen Ersatz, sondern viel mehr eine ntzliche Alternative zum Bayesschen Ansatz dar.

3.4 Weiterentwicklung der Methode

Die in dieser Arbeit entwickelte Methode der latenten Klassenanalyse ist fur die Evaluation bedingt abhangiger diagnostischer Tests ohne Goldstandard anwendbar. Sie benotigt gut gewahlte Startwerte fur alle Parameter und ist damit in der Lage, die wahren unbekannten Parameterwerte zu schatzen. Jedoch gibt es einige Moglichkeiten, die Schatzung zu verbessern und das Anwendungsgebiet zu erweitern.

- Da es viele Anwendungsgebiete der LCA fur bedingt abhangige Einflussvariablen mit mehr als zwei Klassen oder zwei Antwortkategorien gibt, sollte es ein Ziel sein, die

Methode auch auf diese Fälle zu verallgemeinern. Beispielsweise kann eine stärkere Differenzierung des Immunstatus in mehrere Kategorien oder die Beschreibung von Verhaltensweisen oder Eigenschaften das Ziel einer Methodenevaluation sein. Dafür würden dann semi-quantitative Testergebnisse oder nominale Antwortmöglichkeiten zur Beschreibung der Verhaltensweisen benötigt werden. Auf diese Weise würde das Anwendungsgebiet der Methode deutlich erweitert. Jedoch erhöht sich durch die Vergrößerung der Zahl der Klassen und der möglichen Antwortkategorien die Zahl der Parameter deutlich. Beispielsweise gibt es bei einem Modell mit drei Klassen und drei Indikatorvariablen mit drei Kategorien bereits 41 Parameter während bei der gleichen Anzahl von Indikatoren im Zwei-Klassen-Modell mit zwei Antwortkategorien nur 15 Parameter geschätzt werden müssen. Da für alle diese Parameter geeignete Startwerte angegeben werden müssen, erhöht sich der Aufwand durch die Verallgemeinerung stark und es muss geprüft werden, ob das Modell in dieser Form noch praktikabel ist.

- Die Hinzunahme eines Strafterms kann die Schätzung des Modells in vielen Fällen verbessern. Mit nur wenigen unspezifischen Vorinformationen (zum Beispiel Testgüten, die immer größer als 50% sind) kann ein Wertebereich festgelegt werden, in dem die Parameter zu erwarten sind. Dadurch können stark abweichende und unrealistische Resultate vermieden werden. Gegenüber der bisher implementierten Parametergrenzen hat dieser Ansatz den Vorteil, dass er keine feste Grenze benötigt. Stattdessen berücksichtigt er die Stärke der Abweichungen vom festgelegten Parameterbereich, so dass geringere Abweichungen weniger bestraft werden als stark abweichende Parameterwerte. Damit wird die Möglichkeit geschaffen, die Likelihood-Funktion so zu modifizieren, dass die Konvergenz zu den wahren Parameterwerten verbessert wird, ohne die Gefahr, die wahren Parameter durch falsche Annahmen als mögliche Werte auszuschließen.
- Der entwickelte Algorithmus ist so programmiert, dass alle Parameter in jedem Iterationsschritt neu berechnet werden. Auch eventuell zu Beginn bekannte Werte werden so neu abgeschätzt. Dies stellt in der Anwendung zunächst kein Hindernis dar, denn die Simulationsstudien in Publikation I zeigen, dass der Algorithmus, falls dies die tatsächlichen Werte sind, auch zu diesen konvergiert. Wird die Methode jedoch, wie geplant, auf eine vergrößerte Anzahl latenter Klassen mit mehr Antwortkategorien erweitert, führt dies zu einem erhöhten Rechenaufwand. Der Algorithmus kann dann durch eine Fixierung einzelner bekannter Parameterwerte optimiert werden. Sind

beispielsweise Tests voneinander bedingt unabhängig, kann die Festlegung der Abhängigkeiten auf null ohne eine Neuberechnung in jedem Iterationsschritt die Anzahl der zu schätzenden Parameter in einigen Anwendungsfällen deutlich reduzieren. Eine solche Anpassung der Methode verringert zum einen die Anzahl der benötigten Startwerte und erleichtert dem Anwender die Durchführung der Methode. Zum anderen kann es die Konvergenz des Algorithmus beschleunigen.

- Die praktische Anwendbarkeit der vorgestellten frequentistischen LCA kann durch die Verwendung einer interaktiven Webanwendung oder einer Grafischen Benutzeroberfläche verbessert werden. Durch die Entwicklung einer Webseite, zum Beispiel mithilfe des R Paketes Shiny, kann eine Oberfläche erstellt werden, auf welcher der Anwender die benötigten Startinformationen direkt in dafür vorgesehenen Felder eintragen kann. Die Ergebnisse werden in einer solchen Anwendung entsprechend über die genutzte Oberfläche übersichtlich zurückgegeben. Dies hat den Vorteil, dass das entwickelte Verfahren auch ohne Kenntnisse der Programmiersprache R unmittelbar genutzt werden kann. Auch wird der Quelltext dann nicht mehr für die Durchführung der Analyse benötigt und es kann von jedem Computer darauf zugegriffen werden.

Somit gibt es verschiedene Ansatzpunkte, um die vorgestellte Methode in Zukunft weiter zu verbessern und ihre praktische Anwendbarkeit zu erleichtern.

3.5 Übergreifende Schlussfolgerungen und Ausblick

Der in dieser Arbeit vorgestellte Ansatz der latenten Klassenanalyse wurde in zwei Publikationen entwickelt und evaluiert. In Publikation I wurde die Methode hergeleitet und ihre Anwendbarkeit mithilfe von Simulationsstudien untersucht, die realistische Anwendungsfälle aus dem Bereich der veterinärmedizinischen Diagnostik repräsentieren. Die Ergebnisse zeigen, dass die Methode in der Lage ist, die wahren Parameterwerte zu bestimmen. In Publikation II konnte die Anwendbarkeit der Methode auf reale Daten gezeigt werden. Dafür wurde ein Datensatz genutzt, der die Ergebnisse von sechs ELISA Tests zum Nachweis von Antikörpern gegen PRRS enthält, und das notwendige Vorgehen zur Festlegung der Startwerte gezeigt. In der übergreifenden Diskussion wurden technische Herausforderungen der Methode analysiert, ihr Anwendungsbereich ergründet, ein Vergleich zu den Bayessche Methoden gezogen und mögliche Erweiterungen der vorgestellten LCA

diskutiert. Anhand dieser Ergebnisse können folgende Schlussfolgerungen zur Anwendbarkeit der Methode gezogen werden:

- Der vorgestellte frequentistische Ansatz der LCA ist grundsätzlich geeignet, um die Testgüten bedingt abhängiger diagnostischer Tests ohne Goldstandard zu evaluieren.
- Er benötigt einige Voraussetzungen in Form von gut gewählten Startwerten und den Ergebnissen von mindestens drei verschiedenen diagnostischen Tests.
- Die Startwerte lassen sich anhand praktischer Werte abschätzen und benötigen somit keine vorherigen Berechnungen.
- Sind diese Annahmen erfüllt, ist der Algorithmus in der Lage, die wahren Parameterwerte zu schätzen, wie die im Rahmen dieser Arbeit durchgeführten Studien gezeigt haben.
- Somit wurde eine Methode zur Evaluation bedingt abhängiger Tests entwickelt, die für Datensätze mit mindestens drei diagnostischen Tests angewendet auf eine einzelne Population einsetzbar ist. Zusätzlich benötigt sie kein umfangreiches Vorwissen in Form von A-priori-Verteilungen. Damit erfüllt sie die in der Zielstellung definierten Anforderungen.
- Folglich stellt der vorgestellte Ansatz eine geeignete Alternative zu den bereits etablierten Modellen der Bayes-Statistik dar.
- Ihre praktische Anwendbarkeit kann in Zukunft durch eine größere Flexibilität in der Zahl der latenten Klassen und Indikatoren, die Hinzunahme eines Strafterms, die Fixierung einzelner Parameter im Algorithmus sowie die Entwicklung einer interaktiven Webanwendung beziehungsweise einer grafischen Benutzeroberfläche verbessert werden, sodass sie für eine größere Gruppe von Anwendern nutzbar wird.

Insgesamt kann die in dieser Arbeit neu entwickelte Methode also zukünftig für die praktische Evaluation diagnostischer Tests ohne Goldstandard verwendet werden und dadurch zu einer Verbesserung der Kenntnisse der diagnostischen Testgüten beitragen.

4 Zusammenfassung

Clara Schoneberg

Abschätzen der diagnostischen Testgenauigkeit ohne Goldstandard bei bedingt abhängigen Tests – methodische Entwicklung von Modellen der latenten Klassenanalyse

Ein diagnostischer Test in der (Veterinär-)Medizin beurteilt anhand festgelegter Parameter, ob bei einem Individuum ein bestimmtes Ereignis (Krankheit, Infektion etc.) vorliegt. Er bildet die Grundlage für die Berechnung epidemiologischer Maßzahlen wie der Prävalenz und damit für die Einschätzung der Häufigkeit des Auftretens des Ereignisses. Jedoch kommt jeder Test in der Praxis zu einem Anteil falscher Ergebnisse. Dieser Anteil muss bekannt sein, um die Prävalenz korrigieren zu können und eine mögliche Verzerrung der Ergebnisse zu verhindern. Wird ein neuer Test entwickelt, muss zunächst diese Fehlerrate und damit ein Wert für Sensitivität und Spezifität bestimmt werden, indem der Test mit einem (goldenen) Standard verglichen wird.

Liegt ein solcher Goldstandard nicht vor, so stellt die latente Klassenanalyse eine geeignete Methode für den Vergleich dar. Sie nutzt die Ergebnisse mehrerer diagnostischer Tests, die auf dieselben Individuen angewendet werden. Anhand der daraus resultierenden Antworten und dem Anteil übereinstimmender Testergebnisse bestimmt die latente Klassenanalyse die latente Klasse, also den unbekannten Ereignisstatus aller beprobt Individuen. Mit diesen Werten berechnet sie sowohl die Prävalenz in der Stichprobe als auch die Testgüten aller verwendeten Tests. Eine wichtige Voraussetzung ist hierbei die bedingte Unabhängigkeit der Tests. Das bedeutet, dass innerhalb einer latenten Klasse ein Test keine Hinweise auf das Ergebnis eines anderen Tests geben darf. Werden Tests mit demselben biologischen Prinzip verwendet, ist diese Bedingung nicht mehr erfüllt.

Deshalb wurde die latente Klassenanalyse in dieser Arbeit so weiterentwickelt, dass keine bedingte Unabhängigkeit mehr benötigt wird. Die Methode erweitert das klassische latente Klassenmodell um einen Term für die bedingte Abhängigkeit der Tests. Die Parameter schätzt sie dann in einem Algorithmus mithilfe frequentistischer Methoden. In dieser Arbeit wurde diese Methode zunächst anhand von Simulationsstudien evaluiert. Diese zeigten, dass die erweiterte latente Klassenanalyse unter der Voraussetzung gut gewählter Startwerte in der Lage war, die Parameter genau abzuschätzen. Mit diesem Wissen wurde die Methode auf

einen realen Datensatz angewendet. Dieser Datensatz enthielt die Ergebnisse von 812 Schweinen, die mit sechs ELISA-Tests auf das Porzine Reproduktive und Respiratorische Syndrom untersucht wurden. Mit geeigneten Startwerten war der neu entwickelte Algorithmus in der Lage, die Prävalenz, die Testgüten und die Abhängigkeiten der sechs Tests zu bestimmen.

Somit konnte in dieser Arbeit anhand der Ergebnisse beider Studien die Anwendbarkeit der neu entwickelten latenten Klassenanalyse auf bedingt abhängige diagnostische Tests belegt werden. Zusätzliche Modifizierungen für eine vereinfachte Nutzbarkeit und einen größeren Anwendungsbereich sind im Rahmen zukünftiger Arbeiten möglich.

5 Summary

Clara Schoneberg

Estimating the diagnostic test accuracy for conditionally dependent tests without a gold standard – methodical development of latent class analysis models

A diagnostic test in (veterinary) medicine uses defined parameters to assess whether a specific event (disease, infection, etc.) is present in an individual. It forms the basis for calculating epidemiological measures such as prevalence and thus for estimating the frequency of occurrence of the event. However, every test comes in practice to a proportion of false results. This proportion must be known in order to be able to correct the prevalence and prevent a possible bias in the results. If a new test is developed, this error rate and therefore a value for its sensitivity and specificity needs be determined first by comparing the test with a (gold) standard.

If such a gold standard is not available, latent class analysis is a suitable method for comparison. It uses the results of several diagnostic tests applied to the same individuals. Based on the resulting response patterns, the latent class analysis determines the latent class, i.e., the unknown event status of all sampled individuals. Applying these values, the method calculates both the prevalence in the sample as well as the test accuracy of all tests used. An important assumption of the latent class analysis is the conditional independence of the tests. This means that within a latent class a test does not give any indication of the result of another test. If tests with the same biological principle are used, this condition is no longer met.

Therefore, in this work the latent class analysis was further developed so that conditional independence is no longer required. The new method extends the classical latent class model by a term for the conditional dependency. It then estimates the parameters in an algorithm using frequentist methods. In this work the method was first evaluated in different simulated scenarios. They showed that the extended latent class analysis was able to estimate the parameters precisely under the condition of well-chosen starting values. With this knowledge, the method was applied to a real-world data set. The data set contained the results of 812 pigs tested for porcine reproductive and respiratory syndrome using six ELISA tests. The newly developed algorithm was able to determine the prevalence, the test accuracy and the dependencies of the six tests using suitable starting values.

Thus, the applicability of the newly developed latent class analysis for conditionally dependent diagnostic tests was proven in this work based on the results of both studies. Additional modifications of the method for a simplified applicability and a larger area of application are possible in the context of future work.

6 Literaturverzeichnis

- Albert, Paul S, Lisa M McShane, Joanna H Shih, and US National Cancer Institute Bladder Tumor Marker Network. 2001. 'Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors', *Biometrics*, 57: 610-19.
- Berger-Schoch, AE, DC Herrmann, G Schares, Norbert Müller, Daniel Bernet, Bruno Gottstein, and CF Frey. 2011. 'Prevalence and genotypes of Toxoplasma gondii in feline faeces (oocysts) and meat from sheep, cattle and pigs in Switzerland', *Veterinary parasitology*, 177: 290-97.
- Berger, James O. 1985. *Statistical decision theory and Bayesian analysis* (Springer).
- Biernacka, Kinga, Katarzyna Podgórska, Anna Tyszka, and Tomasz Stadejek. 2018. 'Comparison of six commercial ELISAs for the detection of antibodies against porcine reproductive and respiratory syndrome virus (PRRSV) in field serum samples', *Research in veterinary science*, 121: 40-45.
- Black, Michael A, and Bruce A Craig. 2002. 'Estimating disease prevalence in the absence of a gold standard', *Statistics in Medicine*, 21: 2653-69.
- Chen, Caiwei, Wenhui Fan, Xiaojuan Jia, Jing Li, Yuhai Bi, and Wenjun Liu. 2013. 'Development of a recombinant N-Gp5c fusion protein-based ELISA for detection of antibodies to porcine reproductive and respiratory syndrome virus', *Journal of virological methods*, 189: 213-20.
- Collins, Linda M, and Stephanie T Lanza. 2009. *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (John Wiley & Sons).
- Collins, Linda M, and Stuart E Wugalter. 1992. 'Latent class models for stage-sequential dynamic latent variables', *Multivariate Behavioral Research*, 27: 131-57.
- Cong, Yanlong, Zhiqiang Huang, Yixue Sun, Wei Ran, Lisai Zhu, Guilian Yang, Xuemei Ding, Zhanqing Yang, Xiao Huang, and Chunfeng Wang. 2013. 'Development and application of a blocking enzyme-linked immunosorbent assay (ELISA) to differentiate antibodies against live and inactivated porcine reproductive and respiratory syndrome virus', *Virology*, 444: 310-16.
- Coulibaly, Nesson Désiré, and K Robert Yameogo. 2000. 'Prevalence and control of zoonotic diseases: collaboration between public health workers and veterinarians in Burkina Faso', *Acta Tropica*, 76: 53-57.
- Dendukuri, Nandini, and Lawrence Joseph. 2001. 'Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests', *Biometrics*, 57: 158-67.
- Dohoo, Ian Robert, Wayne Martin, and Henrik Stryhn. 2009. *Veterinary epidemiologic research* (AVC Incorporated Charlottetown, Prince Edward Island, Canada).
- Finch, W Holmes, and Kendall Cotton Bronk. 2011. 'Conducting confirmatory latent class analysis using M plus', *Structural Equation Modeling*, 18: 132-51.

- Gardner, Ian A, Axel Colling, and Matthias Greiner. 2019. 'Design, statistical analysis and reporting standards for test accuracy studies for infectious diseases in animals: progress, challenges and recommendations', *Preventive Veterinary Medicine*, 162: 46-55.
- Georgiadis, Marios P, Wesley O Johnson, Ian A Gardner, and Ramanpreet Singh. 2003. 'Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests', *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52: 63-76.
- Greiner, Matthias. 2003. 'Diagnostische Tests in der Veterinärmedizin.' in, *Serodiagnostische Tests* (Springer).
- Hanson, Timothy E, Wesley O Johnson, and Ian A Gardner. 2000. 'Log-linear and logistic modeling of dependence among diagnostic tests', *Preventive Veterinary Medicine*, 45: 123-37.
- Held, Leonhard. 2008. *Methoden der statistischen Inferenz: Likelihood und Bayes* (Spektrum Akademischer Verlag).
- Henao-Diaz, Alexandra, Luis Giménez-Lirola, Ronaldo Magtoto, Ju Ji, and Jeffrey Zimmerman. 2019. 'Evaluation of three commercial porcine reproductive and respiratory syndrome virus (PRRSV) oral fluid antibody ELISAs using samples of known status', *Research in veterinary science*, 125: 113-18.
- Karniychuk, Uladzimir, and Hans Nauwynck. 2014. 'Performance of assays for testing antibodies against porcine reproductive and respiratory syndrome virus in sera collected from swine farms in a region with an extreme virus heterogeneity', *Acta Veterinaria*, 64: 44-51.
- Lanza, Stephanie T, Linda M Collins, David R Lemmon, and Joseph L Schafer. 2007. 'PROC LCA: A SAS procedure for latent class analysis', *Structural equation modeling: a multidisciplinary journal*, 14: 671-94.
- Li, Michael Y. 2018. *An introduction to mathematical modeling of infectious diseases* (Springer).
- Liu, Wei, Bo Zhang, Zhiwei Zhang, Baojiang Chen, and Xiao-Hua Zhou. 2015. 'A pseudo-likelihood approach for estimating diagnostic accuracy of multiple binary medical tests', *Computational Statistics & Data Analysis*, 84: 85-98.
- Lunn, David, David Spiegelhalter, Andrew Thomas, and Nicky Best. 2009. 'The BUGS project: Evolution, critique and future directions', *Statistics in Medicine*, 28: 3049-67.
- Markey, Bryan, Finola Leonard, Marie Archambault, Ann Cullinane, and Dores Maguire. 2013. *Clinical veterinary microbiology e-book* (Elsevier Health Sciences).
- McAlloon, Conor G, Michael L Doherty, Paul Whyte, Cristobal Verdugo, Nils Toft, Simon J More, Luke O'Grady, and Martin J Green. 2019. 'Low accuracy of Bayesian latent class analysis for estimation of herd-level true prevalence under certain disease characteristics—An analysis using simulated data', *Preventive Veterinary Medicine*, 162: 117-25.

- Nielsen, Søren Saxmose, and Nils Toft. 2009. 'A review of prevalences of paratuberculosis in farmed animals in Europe', *Preventive Veterinary Medicine*, 88: 1-14.
- Nylund-Gibson, Karen, and Andrew Young Choi. 2018. 'Ten frequently asked questions about latent class analysis', *Translational Issues in Psychological Science*, 4: 440.
- Parija, Subhash Chandra. 2013. *Textbook of Microbiology & Immunology-E-book* (Elsevier Health Sciences).
- Pouillot, Régis, Guillaume Gerbier, and Ian A Gardner. 2002. "TAGS", a program for the evaluation of test accuracy in the absence of a gold standard', *Preventive Veterinary Medicine*, 53: 67-81.
- Qu, Yinsheng, Ming Tan, and Michael H Kutner. 1996. 'Random effects models in latent class analysis for evaluating accuracy of diagnostic tests', *Biometrics*: 797-810.
- R Development Core Team. 2019. "R: A language and environment for statistical computing." In. Vienna, Austria: R Foundation for Statistical Computing.
- Rogan, Walter J, and Beth Gladen. 1978. 'Estimating prevalence from the results of a screening test', *American journal of epidemiology*, 107: 71-76.
- Sattler, Tatjana, Jutta Pikalo, Eveline Wodak, and Friedrich Schmoll. 2015. 'Performance of ELISAs for detection of antibodies against porcine respiratory and reproductive syndrome virus in serum of pigs after PRRSV type 2 live vaccination and challenge', *Porcine Health Management*, 1: 19.
- _____. 2016. 'Ability of ELISAs to detect antibodies against porcine respiratory and reproductive syndrome virus in serum of pigs after inactivated vaccination and subsequent challenge', *BMC veterinary research*, 12: 259.
- Sattler, Tatjana, Eveline Wodak, Sandra Revilla-Fernández, and Friedrich Schmoll. 2014. 'Comparison of different commercial ELISAs for detection of antibodies against porcine respiratory and reproductive syndrome virus in serum', *BMC veterinary research*, 10: 300.
- Sipos, W, Ch Lang, A Minani, L Fischer, M Ritzmann, and F Schmoll. 2009. 'Comparison of two commercial ELISA systems for the detection of PRRSV-specific antibodies with a gold standard ELISA', *Wien Tierärztl Mschr-Vet Med Austria*, 96: 28-33.
- Tueller, Stephen J, Scott Drotar, and Gitta H Lubke. 2011. 'Addressing the problem of switched class labels in latent variable mixture model simulation studies', *Structural Equation Modeling*, 18: 110-31.
- Vacek, Pamela M. 1985. 'The effect of conditional dependence on the evaluation of diagnostic tests', *Biometrics*, 41: 959-68.
- Virella, Gabriel. 2001. *Medical immunology* (CRC Press).
- Walter, Steven D, and Les M Irwig. 1988. 'Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review', *Journal of clinical epidemiology*, 41: 923-37.

- Wolf, Annika, TL Prüfer, Clara Schoneberg, Amely Campe, Martin Runge, Martin Ganter, and BU Bauer. 2020. 'Prevalence of *Coxiella burnetii* in German sheep flocks and evaluation of a novel approach to detect an infection via preputial swabs at herd-level', *Epidemiology & Infection*, 148.
- World Organisation for Animal Health. 2018. "Manual of diagnostic tests and vaccines for terrestrial animals 2019." In.: OIE Paris, France.
- Wurpts, Ingrid C, and Christian Geiser. 2014. 'Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study', *Frontiers in psychology*, 5: 920.
- Yang, Ilsoon, and Mark P Becker. 1997. 'Latent variable modeling of diagnostic accuracy', *Biometrics*: 948-58.
- Zhang, Bo, Zhen Chen, and Paul S Albert. 2012. 'Estimating diagnostic accuracy of raters without a gold standard by exploiting a group of experts', *Biometrics*, 68: 1294-302.

Anhänge

Anhang 1:

Der folgende Anhang entspricht dem Supplementary Material der Publikation I (Schoneberg C, Kreienbrock L and Campe A (2021) An Iterative, Frequentist Approach for Latent Class Analysis to Evaluate Conditionally Dependent Diagnostic Tests. *Front. Vet. Sci.* 8:588176. doi: 10.3389/fvets.2021.588176).

Abbreviations:

Sen 1 - Sen 3: The sensitivities of tests 1-3

Spe 1 - Spe 3: The specificities of tests 1-3

Prev: The prevalence of the disease in the population

std.: The standardized value of the dependency term

lCl: lower confidence limit

uCl: upper confidence limit

LL: The log-likelihood value for the chosen parameter values

E: Entropy

η_{ij}^+ : The dependency of the sensitivities of test i and test j

η_{ij}^- : The dependency of the specificities of test i and test j

Table S1: Well-chosen starting values for the stepwise latent class algorithm for all five simulation scenarios

Parameter	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Three independent tests						
Prevalence	0.300	0.300	0.300	0.350	0.400	0.400
Sensitivity 1	0.900	0.900	0.900	0.900	0.920	0.920
Sensitivity 2	0.850	0.850	0.850	0.900	0.830	0.830
Sensitivity 3	0.900	0.900	0.900	0.900	0.940	0.940
Specificity 1	0.950	0.950	0.950	0.950	0.920	0.920
Specificity 2	0.950	0.950	0.950	0.950	0.930	0.930
Specificity 3	0.990	0.990	0.990	0.950	0.990	0.990
η_{12}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.010 (0.093)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{13}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.010 (0.111)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{23}^+ (std.)	0.000 (0.000)	0.030 (0.280)	0.010 (0.093)	0.000 (0.000)	0.000 (0.000)	0.030 (0.466)
η_{123}^+ (std.)	0.000 (0.000)	0.001 (0.031)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.001 (0.041)
η_{12}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.003 (0.063)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{13}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.003 (0.138)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{23}^- (std.)	0.000 (0.000)	0.009 (0.350)	0.003 (0.138)	0.000 (0.000)	0.000 (0.000)	0.009 (0.355)

η_{123}^- (std.)	0.000 (0.000)	0.001 (0.100)	-0.003 (-0.635)	0.000 (0.000)	0.000 (0.000)	0.001 (0.145)
Two highly dependent tests with high prevalence						
Prevalence	0.400	0.400	0.400	0.300	0.450	0.450
Sensitivity 1	0.900	0.900	0.900	0.850	0.930	0.930
Sensitivity 2	0.700	0.700	0.700	0.750	0.660	0.660
Sensitivity 3	0.650	0.650	0.650	0.720	0.620	0.620
Specificity 1	0.990	0.990	0.990	0.990	0.990	0.990
Specificity 2	0.800	0.800	0.800	0.850	0.750	0.750
Specificity 3	0.850	0.850	0.850	0.890	0.890	0.890
η_{12}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{13}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{23}^+ (std.)	0.121 (0.600)	0.140 (0.650)	0.090 (0.411)	0.117 (0.600)	0.138 (0.600)	0.110 (0.478)
η_{123}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{12}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{13}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{23}^- (std.)	0.086 (0.600)	0.095 (0.670)	0.065 (0.455)	0.067 (0.600)	0.108 (0.600)	0.095 (0.701)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Two highly dependent tests with low prevalence						
Prevalence	0.030	0.030	0.030	0.010	0.040	0.030
Sensitivity 1	0.900	0.900	0.900	0.850	0.930	0.800
Sensitivity 2	0.700	0.700	0.700	0.750	0.660	0.800
Sensitivity 3	0.650	0.650	0.650	0.720	0.620	0.800
Specificity 1	0.990	0.990	0.990	0.990	0.990	0.800
Specificity 2	0.800	0.800	0.800	0.850	0.750	0.800
Specificity 3	0.850	0.850	0.850	0.890	0.780	0.800
η_{12}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{13}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{23}^+ (std.)	0.121 (0.600)	0.140 (0.650)	0.090 (0.411)	0.117 (0.600)	0.138 (0.600)	0.110 (0.478)
η_{123}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{12}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{13}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{23}^- (std.)	0.086 (0.600)	0.095 (0.670)	0.065 (0.455)	0.067 (0.600)	0.108 (0.600)	0.095 (0.701)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Three medium dependent tests with medium prevalence						
Prevalence	0.400	0.400	0.400	0.350	0.320	0.400

Sensitivity 1	0.800	0.800	0.800	0.850	0.600	0.800
Sensitivity 2	0.660	0.660	0.660	0.730	0.600	0.800
Sensitivity 3	0.700	0.700	0.700	0.750	0.600	0.800
Specificity 1	0.950	0.950	0.950	0.980	0.800	0.800
Specificity 2	0.850	0.850	0.850	0.880	0.800	0.800
Specificity 3	0.880	0.880	0.880	0.920	0.800	0.800
η_{12}^+ (std.)	0.038 (0.200)	0.048 (0.253)	0.025 (0.132)	0.0317 (0.200)	0.048 (0.200)	0.025 (0.156)
η_{13}^+ (std.)	0.046 (0.250)	0.055 (0.300)	0.033 (0.180)	0.0387 (0.250)	0.060 (0.250)	0.033 (0.200)
η_{23}^+ (std.)	0.087 (0.400)	0.097 (0.450)	0.062 (0.286)	0.0769 (0.400)	0.096 (0.400)	0.062 (0.380)
η_{123}^+ (std.)	-0.004 (-0.050)	-0.005 (-0.060)	-0.003 (-0.035)	-0.003 (-0.050)	-0.006 (-0.050)	-0.003 (-0.046)
η_{12}^- (std.)	0.0156 (0.200)	0.020 (0.260)	0.010 (0.128)	0.009 (0.200)	0.032 (0.200)	0.010 (0.063)
η_{13}^- (std.)	0.018 (0.250)	0.022 (0.311)	0.012 (0.170)	0.010 (0.250)	0.040 (0.250)	0.012 (0.075)
η_{23}^- (std.)	0.046 (0.400)	0.050 (0.450)	0.032 (0.280)	0.035 (0.400)	0.064 (0.400)	0.032 (0.200)
η_{123}^- (std.)	-0.0013 (-0.050)	-0.002 (-0.080)	-0.001 (-0.040)	-0.001 (-0.050)	-0.003 (-0.050)	-0.001 (-0.016)
Brucellosis simulation						
Prevalence	0.200	0.200	0.200	0.150	0.200	0.230
Sensitivity 1	0.720	0.720	0.720	0.770	0.650	0.700
Sensitivity 2	0.650	0.650	0.650	0.700	0.650	0.690
Sensitivity 3	0.970	0.970	0.970	0.980	0.950	0.950
Specificity 1	0.980	0.980	0.980	0.990	0.950	0.990
Specificity 2	0.990	0.990	0.990	0.999	0.950	0.990
Specificity 3	0.980	0.980	0.980	0.990	0.950	0.990
η_{12}^+ (std.)	0.129 (0.600)	0.150 (0.700)	0.090 (0.420)	0.116 (0.600)	0.137 (0.600)	0.080 (0.380)
η_{13}^+ (std.)	0.008 (0.100)	0.009 (0.120)	0.003 (0.040)	0.006 (0.100)	0.010 (0.100)	0.000 (0.000)
η_{23}^+ (std.)	0.012 (0.150)	0.015 (0.185)	0.007 (0.086)	0.022 (0.150)	0.0156 (0.150)	0.000 (0.000)
η_{123}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{12}^- (std.)	0.001 (0.100)	0.002 (0.150)	0.001 (0.100)	0.000 ¹ (0.100)	0.005 (0.100)	0.001 (0.100)
η_{13}^- (std.)	0.003 (0.150)	0.004 (0.200)	0.000 (0.000)	0.001 (0.150)	0.007 (0.150)	0.000 (0.000)
η_{23}^- (std.)	0.001 (0.100)	0.002 (0.150)	0.000 (0.000)	0.000 ¹ (0.100)	0.005 (0.100)	0.000 (0.000)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

¹by rounding, η_{ij}^+ : The dependency of the sensitivities of test i and test j, η_{ij}^- : The dependency of the specificities of test i and test j

Table S2: Poorly chosen starting values for the stepwise latent class algorithm for all five simulation scenarios

Parameter	Starting values 1	Starting values 2	Starting values 3
Three independent tests			
Prevalence	0.500	0.250	0.280
Sensitivity 1	0.500	0.850	0.700
Sensitivity 2	0.500	0.850	0.700
Sensitivity 3	0.500	0.850	0.990
Specificity 1	0.500	0.900	0.800
Specificity 2	0.500	0.950	0.800
Specificity 3	0.500	0.930	0.990
η_{12}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.040 (0.200)
η_{13}^+ (std.)	0.000 (0.000)	0.010 (0.080)	0.000 (0.000)
η_{23}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{123}^+ (std.)	0.000 (0.000)	0.000 (0.000)	-0.001 (-0.050)
η_{12}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.009 (0.060)
η_{13}^- (std.)	0.000 (0.000)	0.001 (0.008)	0.000 (0.000)
η_{23}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	-0.001 (-0.063)
Two highly dependent tests with high prevalence			
Prevalence	0.500	0.300	0.395
Sensitivity 1	0.500	0.900	0.696
Sensitivity 2	0.500	0.800	0.895
Sensitivity 3	0.500	0.800	0.845
Specificity 1	0.500	0.990	0.850
Specificity 2	0.500	0.900	0.923
Specificity 3	0.500	0.900	0.973
η_{12}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.010 (0.071)
η_{13}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.010 (0.060)
η_{23}^+ (std.)	0.000 (0.000)	0.010 (0.063)	0.100 (0.090)
η_{123}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{12}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.001 (0.011)
η_{13}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.001 (0.017)
η_{23}^- (std.)	0.000 (0.000)	0.001 (0.011)	0.090 (2.010)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Two highly dependent tests with low prevalence			
Prevalence	0.500	0.040	0.020
Sensitivity 1	0.500	0.900	0.600
Sensitivity 2	0.500	0.800	0.896
Sensitivity 3	0.500	0.800	0.843
Specificity 1	0.500	0.990	0.982
Specificity 2	0.500	0.900	0.924
Specificity 3	0.500	0.900	0.973
η_{12}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.010 (0.071)
η_{13}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.010 (0.060)
η_{23}^+ (std.)	0.000 (0.000)	0.010 (0.063)	0.100 (0.090)
η_{123}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{12}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.001 (0.011)
η_{13}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.001 (0.017)
η_{23}^- (std.)	0.000 (0.000)	0.001 (0.011)	0.090 (2.010)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Three medium dependent tests with medium prevalence			

Prevalence	0.500	0.300	0.375
Sensitivity 1	0.500	0.600	0.790
Sensitivity 2	0.500	0.600	0.795
Sensitivity 3	0.500	0.800	0.857
Specificity 1	0.500	0.900	0.914
Specificity 2	0.500	0.900	0.910
Specificity 3	0.500	0.900	0.951
η_{12}^+ (std.)	0.000 (0.000)	0.010 (0.041)	0.050 (0.300)
η_{13}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.050 (0.350)
η_{23}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.050 (0.354)
η_{123}^+ (std.)	0.000 (0.000)	0.001 (0.010)	0.000 (0.000)
η_{12}^- (std.)	0.000 (0.000)	0.001 (0.011)	0.030 (0.374)
η_{13}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.030 (0.500)
η_{23}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.030 (0.490)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Brucellosis simulation			
Prevalence	0.500	0.300	0.159
Sensitivity 1	0.500	0.850	0.916
Sensitivity 2	0.500	0.800	0.830
Sensitivity 3	0.500	0.700	0.984
Specificity 1	0.500	0.990	0.982
Specificity 2	0.500	0.990	0.992
Specificity 3	0.500	0.900	0.936
η_{12}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.050 (0.500)
η_{13}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.050 (1.500)
η_{23}^+ (std.)	0.000 (0.000)	0.080 (0.440)	0.050 (1.000)
η_{123}^+ (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
η_{12}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.003 (0.253)
η_{13}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.003 (0.092)
η_{23}^- (std.)	0.000 (0.000)	0.001 (0.034)	0.003 (0.150)
η_{123}^- (std.)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)

η_{ij}^+ : The dependency of the sensitivities of test i and test j, η_{ij}^- : The dependency of the specificities of test i and test j

Table S3: Results of the classical and the iterative LCA for the well-chosen starting values in the first simulation scenario

Parameter	LCA for independent test	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.299	0.300	0.301	0.311	0.299	0.300	0.300
lCl	0.271	0.272	0.273	0.282	0.270	0.272	0.271
uCl	0.380	0.390	0.330	0.339	0.327	0.329	0.328
Sen 1	0.910	0.898	0.930	0.875	0.905	0.898	0.937
lCl	0.883	0.879	0.914	0.854	0.887	0.879	0.922
uCl	0.90	0.960	0.946	0.895	0.923	0.916	0.952
Sen 2	0.853	0.850	0.827	0.827	0.858	0.850	0.823
lCl	0.831	0.828	0.803	0.804	0.837	0.828	0.799
uCl	0.885	0.882	0.851	0.850	0.880	0.872	0.846
Sen 3	0.889	0.869	0.892	0.853	0.868	0.889	0.892
lCl	0.908	0.873	0.831	0.847	0.869	0.874	0.834

uCl		0.920	0.875	0.889	0.908	0.912	0.878
Spe 1	0.950	0.949	0.965	0.952	0.950	0.949	0.966
lCl	0.937	0.936	0.953	0.934	0.937	0.935	0.954
uCl	0.964	0.963	0.976	0.965	0.964	0.963	0.977
Spe 2	0.960	0.947	0.959	0.950	0.961	0.960	0.959
lCl	0.982	0.947	0.937	0.949	0.948	0.947	0.933
uCl		0.971	0.964	0.973	0.973	0.971	0.961
Spe 3	0.979	0.981	0.966	0.984	0.978	0.982	0.965
lCl	0.971	0.973	0.954	0.976	0.969	0.973	0.954
uCl	0.988	0.990	0.977	0.992	0.987	0.990	0.976
η_{12}^+ (std.)	-	0.003 (0.048)	-0.004 (-0.040)	0.010 (0.078)	-0.008 (-0.081)	0.003 (0.031)	-0.005 (-0.057)
η_{13}^+ (std.)	-	0.000 ² (0.016)	0.000 ² (-0.002)	0.007 (0.062)	-0.006 (-0.063)	0.000 ² (0.001)	-0.004 (-0.056)
η_{23}^+ (std.)	-	0.000 ² (0.015)	0.025 (0.188)	0.007 (0.054)	-0.007 (-0.059)	0.000 ² (-0.001)	0.024 (0.183)
η_{123}^+ (std.)	-	-0.001 (-0.079)	0.003 (0.099)	0.002 (0.047)	0.007 (0.218)	-0.001 (-0.031)	0.005 (0.160)
η_{12}^- (std.)	-	0.001 (0.002)	0.000 ² (0.006)	0.004 (0.102)	0.002 (0.041)	0.001 (0.013)	0.002 (0.041)
η_{13}^- (std.)	-	-0.001 (-0.040)	0.002 (0.052)	0.003 (0.124)	0.003 (0.081)	-0.001 (-0.019)	0.002 (0.052)
η_{23}^- (std.)	-	-0.001 (-0.041)	0.009 (0.239)	0.003 (0.134)	0.002 (0.078)	-0.001 (-0.024)	0.012 (0.287)
η_{123}^- (std.)	-	0.000 ² (0.166)	0.000 ² (0.010)	-0.003 (-0.737)	-0.003 (-0.470)	0.000 ² (0.045)	-0.001 (-0.167)
It. ¹	-	6	6	6	7	6	7
LL	-1215.97	-1215.97	-1215.97	-1215.97	-1215.97	-1215.97	-1215.97
AIC	14.00	14.00	14.00	14.00	14.00	14.00	14.00
BIC	48.35	48.35	48.35	48.35	48.35	48.35	48.35
E	0.91	0.91	0.91	0.90	0.91	0.91	0.91

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S4: Results of the classical and the iterative LCA for the well-chosen starting values in the second simulation scenario

Parameter	LCA for independent test	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.395	0.400	0.392	0.396	0.421	0.374	0.392
lCl	0.386	0.390	0.383	0.386	0.411	0.365	0.382
uCl	0.405	0.410	0.402	0.405	0.431	0.384	0.401
Sen 1	0.696	0.900	0.930	0.823	0.838	0.977	0.923
lCl	0.687	0.894	0.925	0.815	0.831	0.975	0.918
uCl	0.705	0.906	0.935	0.830	0.845	0.981	0.928
Sen 2	0.895	0.700	0.693	0.758	0.713	0.691	0.697
lCl	0.889	0.691	0.684	0.750	0.704	0.682	0.688
uCl	0.901	0.709	0.702	0.767	0.721	0.700	0.706
Sen 3	0.845	0.650	0.643	0.708	0.662	0.642	0.646
lCl	0.838	0.641	0.633	0.699	0.653	0.632	0.637
uCl	0.852	0.659	0.652	0.717	0.672	0.651	0.655
Spe 1	0.850	0.990	0.998	0.933	0.977	0.999	0.992

lCl	0.843	0.988	0.997	0.928	0.974	0.998	0.990
uCl	0.857	0.992	0.999	0.938	0.980	1.000	0.994
Spe 2	0.923	0.800	0.789	0.834	0.827	0.774	0.791
lCl	0.918	0.792	0.780	0.827	0.820	0.766	0.783
uCl	0.929	0.808	0.797	0.842	0.834	0.783	0.799
Spe 3	0.973	0.850	0.839	0.884	0.877	0.824	0.841
lCl	0.970	0.843	0.832	0.878	0.870	0.817	0.833
uCl	0.977	0.857	0.846	0.890	0.883	0.832	0.848
η_{12}^+ (std.)	-	0.000 (0.000)	0.001 (0.006)	0.000 ² (0.002)	-0.001 (-0.008)	0.000 ² (0.006)	0.001 (0.010)
η_{13}^+ (std.)	-	0.000 (0.000)	0.001 (0.004)	0.000 ² (0.002)	-0.001 (-0.006)	0.000 ² (0.004)	0.001 (0.011)
η_{23}^+ (std.)	-	0.121 (0.600)	0.125 (0.567)	0.093 (0.477)	0.116 (0.544)	0.125 (0.564)	0.121 (0.552)
η_{123}^+ (std.)	-	0.000 (0.000)	-0.001 (-0.014)	0.000 ² (-0.002)	0.000 ² (-0.002)	0.000 ² (-0.003)	0.000 ² (0.006)
η_{12}^- (std.)	-	0.000 (0.000)	0.000 ² (-0.003)	-0.001 (-0.006)	0.000 ² (0.004)	0.000 ² (-0.083)	0.000 (0.000)
η_{13}^- (std.)	-	0.000 (0.000)	0.000 ² (-0.003)	0.000 ² (-0.006)	0.000 ² (0.005)	0.000 ² (-0.079)	0.000 ² (0.006)
η_{23}^- (std.)	-	0.086 (0.600)	0.091 (0.609)	0.065 (0.547)	0.069 (0.557)	0.100 (0.627)	0.092 (0.617)
η_{123}^- (std.)	-	0.000 (0.000)	0.000 ² (0.024)	0.000 ² (0.006)	0.000 ² (-0.020)	0.000 ² (0.325)	0.000 ² (-0.029)
It. ¹	-	3	4	6	5	4	5
LL	-15968.1	-15968.1	-15968.1	-15968.1	-15968.1	-15968.1	-15968.1
AIC	14.00	14.00	14.00	14.00	14.00	14.00	14.00
BIC	64.47	64.47	64.47	64.47	64.47	64.47	64.47
E	0.826	0.78	0.82	0.71	0.73	0.90	0.80

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S5: Results of the classical and the iterative LCA for the well-chosen starting values in the third simulation scenario

Parameter	LCA for independent test	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.169	0.030	0.031	0.071	0.024	0.0002	0.029
lCl	0.161	0.027	0.028	0.066	0.021	0.0000	0.025
uCl	0.177	0.033	0.034	0.076	0.027	0.0006	0.032
Sen 1	0.130	0.900	0.912	0.328	0.859	0.036	0.888
lCl	0.124	0.894	0.906	0.319	0.852	0.032	0.882
uCl	0.137	0.906	0.918	0.337	0.866	0.040	0.894
Sen 2	0.896	0.701	0.672	0.799	0.836	0.670	0.726
lCl	0.890	0.692	0.662	0.791	0.828	0.661	0.717
uCl	0.902	0.710	0.681	0.807	0.843	0.679	0.735
Sen 3	0.843	0.650	0.621	0.748	0.766	1.000	0.676
lCl	0.835	0.640	0.612	0.739	0.758	NaN	0.667
uCl	0.850	0.659	0.631	0.756	0.774	NaN	0.685
Spe 1	0.982	0.990	0.991	0.986	0.984	0.963	0.988
lCl	0.979	0.988	0.990	0.983	0.981	0.960	0.986
uCl	0.985	0.992	0.993	0.988	0.986	0.967	0.991

Spe 2	0.924	0.800	0.800	0.830	0.800	0.785	0.800
ICl	0.919	0.792	0.792	0.822	0.792	0.777	0.792
uCl	0.929	0.808	0.808	0.837	0.808	0.793	0.808
Spe 3	0.973	0.850	0.850	0.880	0.850	0.835	0.850
ICl	0.970	0.843	0.843	0.873	0.843	0.828	0.667
uCl	0.976	0.857	0.857	0.886	0.857	0.843	0.685
η_{12}^+ (std.)	-	0.000 (0.000)	0.001 (0.005)	-0.005 (-0.027)	-0.009 (-0.066)	-0.078 (-0.892)	0.001 (0.010)
η_{13}^+ (std.)	-	0.000 (0.000)	0.000 ² (-0.001)	-0.005 (-0.024)	-0.018 (-0.120)	-0.049 (NaN)	0.000 (0.000)
η_{23}^+ (std.)	-	0.121 (0.600)	0.129 (0.565)	0.085 (0.487)	0.104 (0.665)	-0.506 (NaN)	0.102 (0.490)
η_{123}^+ (std.)	-	0.000 (0.000)	0.002 (0.037)	-0.001 (-0.014)	0.002 (0.036)	0.099 (NaN)	0.003 (0.052)
η_{12}^- (std.)	-	0.000 (0.000)	0.000 ² (0.003)	0.000 ² (0.007)	0.001 (0.013)	0.013 (0.169)	0.000 ² (0.002)
η_{13}^- (std.)	-	0.000 (0.000)	0.000 ² (0.003)	0.000 ² (0.006)	0.001 (0.027)	0.013 (0.186)	0.000 ² (0.002)
η_{23}^- (std.)	-	0.086 (0.600)	0.086 (0.601)	0.067 (0.545)	0.084 (0.590)	0.094 (0.617)	0.086 (0.601)
η_{123}^- (std.)	-	0.000 (0.000)	0.000 ² (-0.008)	0.000 ² (-0.008)	0.000 (0.000)	-0.007 (-0.245)	0.000 ² (-0.008)
It. ¹	-	4	5	9	6	4	5
LL	-9522.50	-9522.50	-9522.50	-9522.50	-9522.50	-9522.50	-9522.50
AIC	14.00	14.00	14.00	14.00	14.00	14.00	14.00
BIC	64.47	64.47	64.47	64.47	64.47	64.47	64.47
E	0.80	0.74	0.74	0.64	0.78	0.61	0.74

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S6: Results of the classical and the iterative LCA for the well-chosen starting values in the fourth simulation scenario

Parameter	LCA for independent test	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.375	0.400	0.402	0.402	0.395	0.345	0.380
ICl	0.374	0.390	0.392	0.393	0.385	0.336	0.370
uCl	0.384	0.410	0.411	0.412	0.405	0.355	0.390
Sen 1	0.790	0.800	0.793	0.791	0.844	0.741	0.805
ICl	0.782	0.792	0.785	0.783	0.837	0.732	0.797
uCl	0.798	0.808	0.801	0.800	0.850	0.749	0.813
Sen 2	0.794	0.660	0.642	0.694	0.666	0.711	0.716
ICl	0.787	0.651	0.633	0.685	0.656	0.702	0.707
uCl	0.803	0.669	0.652	0.703	0.675	0.720	0.725
Sen3	0.857	0.700	0.684	0.733	0.706	0.737	0.759
ICl	0.850	0.690	0.675	0.724	0.697	0.728	0.750
uCl	0.864	0.710	0.694	0.742	0.715	0.746	0.768
Spe 1	0.914	0.950	0.947	0.950	0.972	0.856	0.929
ICl	0.908	0.946	0.942	0.943	0.969	0.849	0.923
uCl	0.919	0.954	0.952	0.952	0.975	0.863	0.934
Spe 2	0.910	0.850	0.840	0.875	0.849	0.834	0.868
ICl	0.905	0.843	0.832	0.868	0.842		0.861

uCl	0.916	0.857	0.847	0.881	0.856	0.827	0.874
Spe 3	0.951	0.880	0.871	0.905	0.880	0.851	0.898
lCl	0.944	0.874	0.865	0.899	0.873	0.844	0.892
uCl	0.955	0.886	0.878	0.910	0.886	0.858	0.904
η_{12}^+ (std.)	-	0.038 (0.200)	0.047 (0.242)	0.024 (0.130)	0.026 (0.154)	0.055 (0.277)	0.023 (0.127)
η_{13}^+ (std.)	-	0.046 (0.250)	0.054 (0.287)	0.033 (0.182)	0.034 (0.201)	0.069 (0.357)	0.031 (0.185)
η_{23}^+ (std.)	-	0.087 (0.400)	0.097 (0.435)	0.066 (0.330)	0.084 (0.400)	0.078 (0.390)	0.058 (0.302)
η_{123}^+ (std.)	-	-0.004 (-0.050)	-0.005 (-0.051)	-0.002 (-0.022)	-0.005 (-0.063)	-0.004 (-0.045)	-0.004 (-0.050)
η_{12}^- (std.)	-	0.016 (0.200)	0.020 (0.238)	0.009 (0.127)	0.009 (0.154)	0.036 (0.277)	0.011 (0.126)
η_{13}^- (std.)	-	0.018 (0.250)	0.021 (0.285)	0.012 (0.178)	0.009 (0.173)	0.046 (0.369)	0.013 (0.164)
η_{23}^- (std.)	-	0.046 (0.400)	0.051 (0.411)	0.033 (0.338)	0.047 (0.400)	0.052 (0.394)	0.034 (0.330)
η_{123}^- (std.)	-	-0.001 (-0.050)	-0.001 (-0.019)	-0.001 (-0.051)	-0.001 (-0.021)	-0.004 (-0.091)	-0.001 (-0.032)
It. ¹	-	3	6	5	6	7	4
LL	-15673.82	-15673.8	-15673.8	-15673.8	-15673.8	-15673.8	-15673.8
AIC	14.00	14.00	14.00	14.00	14.00	14.00	14.00
BIC	64.47	64.47	64.47	64.47	64.47	64.47	64.47
E	0.80	0.72	0.70	0.74	0.76	0.67	0.80

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

Table S7: Results of the classical and the iterative LCA for the well-chosen starting values in the fifth simulation scenario

Parameter	LCA for independent test	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0159	0.200	0.208	0.187	0.204	0.186	0.180
lCl	0.151	0.192	0.200	0.179	0.196	0.178	0.173
uCl	0.166	0.208	0.216	0.195	0.211	0.194	0.188
Sen 1	0.916	0.720	0.691	0.781	0.729	0.748	0.801
lCl	0.910	0.711	0.682	0.773	0.720	0.740	0.793
uCl	0.921	0.729	0.700	0.789	0.737	0.757	0.809
Sen 2	0.830	0.650	0.622	0.706	0.645	0.684	0.725
lCl	0.822	0.641	0.613	0.700	0.635	0.675	0.716
uCl	0.837	0.659	0.632	0.714	0.654	0.693	0.733
Sen3	0.984	0.970	0.964	0.973	0.953	0.961	0.991
lCl	0.982	0.967	0.960	0.970	0.948	0.958	0.990
uCl	0.987	0.973	0.968	0.976	0.957	0.965	0.993
Spe 1	0.982	0.980	0.979	0.983	0.986	0.974	0.981
lCl	0.979	0.977	0.977	0.980	0.983	0.971	0.979
uCl	0.985	0.983	0.982	0.986	0.988	0.978	0.984
Spe 2	0.992	0.990	0.989	0.992	0.992	0.987	0.991
lCl	0.990	0.988	0.987	0.990	0.990	0.985	0.989
uCl	0.994	0.992	0.991	0.994	0.994	0.989	0.993
Spe 3	0.935	0.980	0.988	0.965	0.980	0.962	0.962

ICl uCl	0.930 0.940	0.977 0.983	0.986 0.990	0.962 0.969	0.977 0.983	0.958 0.965	0.958 0.966
η_{12}^+ (std.)	-	0.129 (0.600)	0.145 (0.649)	0.094 (0.500)	0.121 (0.568)	0.127 (0.628)	0.084 (0.470)
η_{13}^+ (std.)	-	0.008 (0.100)	0.011 (0.126)	0.005 (0.071)	0.007 (0.072)	0.015 (0.176)	0.002 (0.042)
η_{23}^+ (std.)	-	0.012 (0.150)	0.017 (0.191)	0.009 (0.120)	0.023 (0.227)	0.016 (0.175)	0.002 (0.036)
η_{123}^+ (std.)	-	0.000 (0.000)	-0.001 (-0.028)	-0.002 (-0.015)	0.000 ² (-0.021)	0.0001 (0.018)	-0.002 (-0.110)
η_{12}^- (std.)	-	0.001 (0.100)	0.001 (0.072)	0.000 (0.000)	0.000 ² (0.020)	0.002 (0.101)	0.001 (0.058)
η_{13}^- (std.)	-	0.003 (0.150)	0.004 (0.241)	0.000 (0.000)	0.000 ² (0.064)	0.008 (0.263)	0.000 ² (-0.015)
η_{23}^- (std.)	-	0.001 (0.100)	0.002 (0.164)	0.000 (0.000)	0.000 ² (0.001)	0.005 (0.234)	0.000 ² (-0.020)
η_{123}^- (std.)	-	0.000 (0.000)	0.001 (0.312)	0.000 (0.000)	0.000 ² (0.169)	0.000 ² (0.038)	0.000 ² (0.169)
It. ¹	-	3	6	7	6	7	6
LL	-8624.43	-8624.43	-8624.43	-8624.43	-8624.43	-8624.43	-8624.43
AIC	14.00	14.00	14.00	14.00	14.00	14.00	14.00
BIC	64.47	64.47	64.47	64.47	64.47	64.47	64.47
E	0.95	0.92	0.92	0.91	0.91	0.91	0.93

¹The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S8: Results of the iterative LCA for the poorly chosen starting values in the first simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (ICl,uCl)	0.500 (0.469,0.531)	0.302 (0.273,0.330)	0.304 (0.275,0.332)
Sen 1 (ICl,uCl)	0.305 (0.276, 0.334)	0.893 (0.874,0.912)	0.862 (0.840,0.883)
Sen 2 (ICl,uCl)	0.284 (0.256,0.312)	0.859 (0.837,0.880)	0.816 (0.791,0.840)
Sen 3 (ICl,uCl)	0.281 (0.253,0.309)	0.880 (0.860,0.900)	0.922 (0.906,0.939)
Spe 1 (ICl,uCl)	0.695 (0.666,0.724)	0.949 (0.935,0.963)	0.938 (0.923,0.953)
Spe 2 (ICl,uCl)	0.716 (0.688,0.744)	0.964 (0.953,0.976)	0.948 (0.934,0.962)
Spe 3 (ICl,uCl)	0.719 (0.691,0.747)	0.978 (0.969,0.987)	0.999 (0.996,1.000)
η_{12}^+ (std.)	0.145 (0.700)	-0.005 (-0.055)	0.029 (0.219)
η_{13}^+ (std.)	0.155 (0.750)	0.004 (0.041)	-0.003 (-0.032)
η_{23}^+ (std.)	0.148 (0.731)	-0.006 (-0.049)	-0.003 (-0.030)
η_{123}^+ (std.)	0.051 (0.541)	0.006 (0.160)	0.002 (0.047)
η_{12}^- (std.)	0.145 (0.700)	0.002 (0.041)	0.011 (0.200)
η_{13}^- (std.)	0.155 (0.750)	0.003 (0.086)	0.001 (0.096)
η_{23}^- (std.)	0.148 (0.731)	0.002 (0.063)	0.001 (0.095)
η_{123}^- (std.)	-0.051 (-0.541)	-0.002 (-0.385)	-0.001 (-0.715)

Iterations ¹	3	7	6
LL	-1215.97	-1215.97	-1215.97
AIC	14.00	14.00	14.00
BIC	48.35	48.35	48.35
E	0.00	0.91	0.93

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

Table S9: Results of the iterative LCA for the poorly chosen starting values in the second simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.500 (0.490,0.510)	0.403 (0.394,0.413)	0.358 (0.349,0.367)
Sen 1 (lCl,uCl)	0.366 (0.357, 0.375)	0.695 (0.686,0.704)	0.900 (0.895,0.907)
Sen 2 (lCl,uCl)	0.400 (0.390,0.410)	0.880 (0.874,0.886)	0.746 (0.737,0.754)
Sen 3 (lCl,uCl)	0.350 (0.341,0.359)	0.828 (0.821,0.836)	0.696 (0.687,0.705)
Spe 1 (lCl,uCl)	0.634 (0.625,0.644)	0.856 (0.849,0.863)	0.932 (0.927,0.937)
Spe 2 (lCl,uCl)	0.600 (0.590,0.610)	0.925 (0.919,0.930)	0.792 (0.785,0.801)
Spe 3 (lCl,uCl)	0.650 (0.641,0.659)	0.973 (0.970,0.977)	0.843 (0.836,0.850)
η_{12}^+ (std.)	0.107 (0.453)	0.002 (0.015)	0.009 (0.065)
η_{13}^+ (std.)	0.107 (0.465)	0.002 (0.015)	0.009 (0.062)
η_{23}^+ (std.)	0.160 (0.685)	0.012 (0.098)	0.096 (0.481)
η_{123}^+ (std.)	0.018 (0.162)	-0.001 (-0.020)	0.001 (0.012)
η_{12}^- (std.)	0.107 (0.453)	0.000 ² (-0.013)	0.001 (0.009)
η_{13}^- (std.)	0.107 (0.465)	0.000 ² (-0.015)	0.001 (0.008)
η_{23}^- (std.)	0.160 (0.685)	0.000 ² (-0.002)	0.091 (0.621)
η_{123}^- (std.)	-0.018 (-0.163)	0.001(0.075)	0.000 ² (0.025)
Iterations ¹	3	7	6
LL	-15968.05	-15968.05	-15968.05
AIC	14.00	14.00	14.00
BIC	64.47	64.47	64.47
E	0.00	0.82	0.71

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

² by rounding

Table S10: Results of the iterative LCA for the poorly chosen starting values in the third simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.500 (0.490,0.510)	0.158 (0.150,0.165)	0.000 (0.000,0.010)
Sen 1 (lCl,uCl)	0.037 (0.033,0.040)	0.137 (0.130,0.143)	0.063 (0.059,0.068)
Sen 2 (lCl,uCl)	0.215 (0.207,0.223)	0.913 (0.908,0.919)	0.364 (0.354,0.373)
Sen 3	0.165	0.866	0.655

(ICl,uCl)	(0.158,0.172)	(0.860,0.873)	(0.646,0.665)
Spe 1 (ICl,uCl)	0.963 (0.960,0.967)	0.982 (0.979,0.985)	0.962 (0.959,0.966)
Spe 2 (ICl,uCl)	0.785 (0.777,0.793)	0.916 (0.910,0.921)	0.780 (0.771,0.788)
Spe 3 (ICl,uCl)	0.835 (0.828,0.842)	0.966 (0.963,0.970)	0.817 (0.809,0.825)
η_{12}^+ (std.)	0.013 (0.168)	-0.004 (-0.043)	0.024 (0.201)
η_{13}^+ (std.)	0.013 (0.185)	-0.004 (-0.038)	-0.029 (-0.253)
η_{23}^+ (std.)	0.094 (0.617)	0.006 (0.056)	0.056 (0.247)
η_{123}^+ (std.)	0.007 (0.245)	0.000 ² (0.001)	-0.002 (-0.513)
η_{12}^- (std.)	0.013 (0.168)	0.001 (0.022)	0.014 (0.171)
η_{13}^- (std.)	0.013 (0.185)	0.001 (0.024)	0.012 (0.161)
η_{23}^- (std.)	0.094 (0.617)	0.002 (0.037)	0.095 (0.594)
η_{123}^- (std.)	-0.007 (-0.245)	-0.001 (-0.109)	-0.007 (-0.218)
Iterations ¹	3	9	9
LL	-9522.50	-9522.50	-9522.50
AIC	14.00	14.00	14.00
BIC	64.47	64.47	64.47
E	0.00	0.811	0.01

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S11: Results of the iterative LCA for the poorly chosen starting values in the fourth simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (ICl,uCl)	0.500 (0.490,0.510)	0.386 (0.377,0.397)	0.053 (0.049,0.058)
Sen 1 (ICl,uCl)	0.350 (0.340,0.360)	0.770 (0.762,0.778)	1.000 (NaN,NaN)
Sen 2 (ICl,uCl)	0.354 (0.345,0.363)	0.774 (0.766,0.783)	0.849 (0.842,0.856)
Sen 3 (ICl,uCl)	0.352 (0.343,0.361)	0.856 (0.849,0.863)	0.202 (194,0.210)
Spe 1 (ICl,uCl)	0.650 (0.64,0.660)	0.914 (0.909,0.919)	0.688 (0.679,0.697)
Spe 2 (ICl,uCl)	0.646 (0.637,0.655)	0.910 (0.905,0.916)	0.674 (0.665,0.683)
Spe 3 (ICl,uCl)	0.648 (0.639,0.657)	0.969 (0.962,0.969)	0.640 (0.630,0.650)
η_{12}^+ (std.)	0.116 (0.510)	0.013 (0.077)	-0.161 (NaN)
η_{13}^+ (std.)	0.133 (0.585)	0.002 (0.014)	-0.192 (NaN)
η_{23}^+ (std.)	0.134 (0.585)	0.002 (0.012)	-0.179 (-1.246)
η_{123}^+ (std.)	0.024 (0.216)	0.000 ² (-0.003)	0.154 (NaN)
η_{12}^- (std.)	0.116 (0.510)	0.000 ² (0.002)	0.112 (0.515)
η_{13}^- (std.)	0.133 (0.585)	-0.001 (-0.020)	0.158 (0.709)
η_{23}^- (std.)	0.134 (0.585)	-0.001 (-0.020)	0.156 (0.691)
η_{123}^- (std.)	-0.024 (-0.216)	0.012 (0.082)	-0.039 (-0.378)
Iterations ¹	3	7	6
LL	-15673.82	-15673.82	-15673.82

AIC	14.00	14.00	14.00
BIC	64.47	64.47	64.47
E	0.00	0.80	0.00

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S12: Results of the iterative LCA for the poorly chosen starting values in the fifth simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.500 (0.490,0.510)	0.173 (0.166,0.180)	0.000 (NaN,NaN)
Sen 1 (lCl,uCl)	0.160 (0.153,0.167)	0.924 (0.919,0.929)	0.473 (0.463,0.483)
Sen 2 (lCl,uCl)	0.138 (0.131,0.145)	0.755 (0.746,0.762)	0.786 (0.778,0.794)
Sen 3 (lCl,uCl)	0.210 (0.202,0.218)	0.900 (0.894,0.906)	0.670 (0.660,0.679)
Spe 1 (lCl,uCl)	0.840 (0.833,0.847)	1.000 (0.999,1.000)	0.840 (0.833,0.847)
Spe 2 (lCl,uCl)	0.862 (0.855,0.869)	0.991 (0.989,883)	0.862 (0.855,0.869)
Spe 3 (lCl,uCl)	0.790 (0.782,0.798)	0.934 (0.929,0.939)	0.790 (0.782,0.798)
η_{12}^+ (std.)	0.098 (0.779)	0.000 ² (-0.002)	0.174 (0.853)
η_{13}^+ (std.)	0.110 (0.738)	0.000 ² (0.003)	0.121 (0.516)
η_{23}^+ (std.)	0.100 (0.718)	0.064 (0.495)	0.410 (2.126)
η_{123}^+ (std.)	0.0612 (1.202)	-0.001 (-0.027)	-0.103 (-1.070)
η_{12}^- (std.)	0.098 (0.779)	0.000 ² (0.025)	0.098 (0.779)
η_{13}^- (std.)	0.110 (0.738)	0.000 ² (0.013)	0.110 (0.738)
η_{23}^- (std.)	0.100 (0.718)	0.001 (0.044)	0.101 (0.718)
η_{123}^- (std.)	-0.062 (-1.202)	0.000 ² (-0.147)	-0.062 (-1.202)
Iterations ¹	3	5	5
LL	-8624.429	-8624.429	-8624.429
AIC	14.00	14.00	14.00
BIC	64.47	64.47	64.47
E	0.00	0.94	0.63

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S13: Results of the classical and the iterative LCA for the well-chosen starting values in the third simulation scenario with only positive pairwise dependencies allowed

Parameter	LCA for independent test	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.169	0.030	0.031	0.071	0.024	0.028	0.029
lCl	0.161	0.027	0.028	0.066	0.021	0.025	0.025
uCl	0.177	0.033	0.034	0.076	0.027	0.031	0.032
Sen 1	0.130	0.900	0.912	0.325	0.857	0.958	0.888
lCl	0.124	0.894	0.906	0.316	0.850	0.953	0.882
uCl	0.137	0.906	0.918	0.334	0.864	0.962	0.894

Sen 2	0.896	0.701	0.672	0.796	0.832	0.688	0.726
ICl	0.890	0.692	0.662	0.788	0.824	0.679	0.717
uCl	0.902	0.710	0.681	0.804	0.839	0.697	0.735
Sen 3	0.843	0.650	0.621	0.745	0.756	0.640	0.676
ICl	0.835	0.640	0.612	0.737	0.749	0.630	0.667
uCl	0.850	0.659	0.631	0.753	0.766	0.649	0.685
Spe 1	0.982	0.990	0.991	0.986	0.983	0.990	0.988
ICl	0.979	0.988	0.990	0.983	0.980	0.988	0.986
uCl	0.985	0.992	0.993	0.988	0.986	0.992	0.991
Spe 2	0.924	0.800	0.800	0.830	0.800	0.800	0.800
ICl	0.919	0.792	0.792	0.822	0.792	0.790	0.792
uCl	0.929	0.808	0.808	0.837	0.808	0.810	0.808
Spe 3	0.973	0.850	0.850	0.880	0.850	0.849	0.850
ICl	0.970	0.843	0.843	0.873	0.843	0.842	0.667
uCl	0.976	0.857	0.857	0.886	0.857	0.856	0.685
η_{12}^+ (std.)	-	0.000 (0.000)	0.000 (0.005)	0.000 (0.000)	0.001 (0.004)	0.000 ² (0.002)	0.001 (0.010)
η_{13}^+ (std.)	-	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 ² (0.002)	0.000 (0.000)
η_{23}^+ (std.)	-	0.121 (0.600)	0.129 (0.565)	0.082 (0.464)	0.102 (0.637)	0.111 (0.500)	0.102 (0.490)
η_{123}^+ (std.)	-	0.000 (0.000)	0.002 (0.037)	-0.001 (-0.004)	0.004 (0.064)	0.009 (0.196)	0.003 (0.052)
η_{12}^- (std.)	-	0.000 (0.000)	0.000 ² (0.003)	0.000 ² (0.003)	0.001 (0.013)	0.000 ² (0.010)	0.000 ² (0.002)
η_{13}^- (std.)	-	0.000 (0.000)	0.000 ² (0.003)	0.000 ² (0.003)	0.001 (0.024)	0.000 ² (0.007)	0.000 ² (0.002)
η_{23}^- (std.)	-	0.086 (0.600)	0.086 (0.601)	0.067 (0.548)	0.085 (0.590)	0.087 (0.604)	0.086 (0.601)
η_{123}^- (std.)	-	0.000 (0.000)	0.000 ² (-0.008)	0.000 ² (0.009)	0.000 ² (0.024)	0.000 ² (-0.023)	0.000 ² (-0.008)
Iterations ¹	-	4	5	9	7	5	7
LL	-9522.50	-9522.50	-9522.50	-9522.50	-9522.50	-9522.50	-9522.50
AIC	14.00	14.00	14.00	14.00	14.00	14.00	14.00
BIC	64.47	64.47	64.47	64.47	64.47	64.47	64.47
E	0.80	0.74	0.74	0.64	0.78	0.80	0.74

¹The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S14: Results of the iterative LCA for the third simulation scenario for the poorly chosen starting values with only positive pairwise dependencies allowed

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (ICl,uCl)	0.500 (0.490,0.510)	0.177 (0.150,0.165)	0.023 (0.020,0.026)
Sen 1 (ICl,uCl)	0.037 (0.033,0.040)	0.121 (0.115,0.128)	0.977 (0.974,0.980)
Sen 2 (ICl,uCl)	0.215 (0.207,0.223)	0.862 (0.856,0.869)	0.759 (0.751,0.768)
Sen 3 (ICl,uCl)	0.165 (0.158,0.172)	0.825 (0.818,0.833)	0.707 (0.698,0.761)
Spe 1	0.963	0.981	0.985

(lCl,uCl)	(0.960,0.967)	(0.978,0.985)	(0.983,0.988)
Spe 2 (lCl,uCl)	0.785 (0.777,0.793)	0.924 (0.919,0.930)	0.797 (0.790,0.806)
Spe 3 (lCl,uCl)	0.835 (0.828,0.842)	0.977 (0.974,0.980)	0.847 (0.841,0.855)
η_{12}^+ (std.)	0.013 (0.168)	0.000 (0.000)	0.011 (0.169)
η_{13}^+ (std.)	0.013 (0.185)	0.000 (0.000)	0.010 (0.153)
η_{23}^+ (std.)	0.094 (0.617)	0.000 (0.000)	0.106 (0.546)
η_{123}^+ (std.)	0.007 (0.245)	0.000 ² (0.005)	-0.001 (-0.024)
η_{12}^- (std.)	0.013 (0.168)	0.002 (0.045)	0.001 (0.018)
η_{13}^- (std.)	0.013 (0.185)	0.001 (0.061)	0.001 (0.020)
η_{23}^- (std.)	0.094 (0.617)	0.003 (0.071)	0.087 (0.601)
η_{123}^- (std.)	-0.007 (-0.245)	-0.002 (-0.289)	-0.001 (-0.04)
Iterations ¹	3	6	4
LL	-9522.498	-9522.498	-9522.498
AIC	14.00	14.00	14.00
BIC	64.47	64.47	64.47
E	0.00	0.77	0.01

¹ The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

²by rounding

Table S15: Results of the Bayesian LCA for the well-chosen starting values in the first simulation scenario

Para-meter	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.300	0.301	0.300	0.338	0.369	0.369
lCl	0.293	0.286	0.285	0.330	0.361	0.361
uCl	0.307	0.316	0.315	0.346	0.377	0.377
Sen 1	0.896	0.882	0.882	0.900	0.912	0.912
lCl	0.887	0.860	0.860	0.810	0.904	0.904
uCl	0.906	0.904	0.904	0.909	0.920	0.920
Sen 2	0.845	0.833	0.832	0.886	0.828	0.828
lCl	0.835	0.811	0.811	0.877	0.818	0.818
uCl	0.855	0.854	0.853	0.895	0.838	0.838
Sen 3	0.900	0.899	0.901	0.899	0.936	0.936
lCl	0.891	0.878	0.879	0.890	0.929	0.929
uCl	0.909	0.920	0.923	0.908	0.943	0.943
Spe 1	0.947	0.941	0.940	0.947	0.925	0.925
lCl	0.941	0.929	0.928	0.941	0.918	0.918
uCl	0.953	0.953	0.952	0.953	0.932	0.932
Spe 2	0.950	0.949	0.947	0.953	0.931	0.931
lCl	0.944	0.938	0.935	0.947	0.924	0.924
uCl	0.956	0.960	0.959	0.960	0.938	0.938
Spe 3	0.989	0.985	0.985	0.989	0.981	0.981
lCl	0.986	0.974	0.973	0.986	0.977	0.977
uCl	0.992	0.996	0.997	0.992	0.985	0.985
η_{12}^+	0.020	0.0223	0.021	0.017	0.019	0.018
η_{13}^+	0.000 ¹	0.008	0.008	0.003	0.004	0.004
η_{23}^+	0.002	0.007	0.008	-0.002	0.011	0.012
η_{123}^+	0.000 ¹	0.000 ¹	0.000 ¹	0.000 ¹	0.019	0.000 ¹

η_{12}^-	0.005	0.009	0.011	0.003	0.004	0.019
η_{13}^-	0.003	0.004	0.004	0.003	0.007	0.003
η_{23}^-	0.004	0.005	0.005	0.003	-0.001	0.007
η_{123}^-	-0.001	0.000 ¹	0.000 ¹	0.000 ¹	-0.001	0.000 ¹
Iterations	10000	10000	10000	10000	10000	10000

¹by rounding

Table S16: Results of the Bayesian LCA for the well-chosen starting values in the second simulation scenario

Para-meter	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.400	0.402	0.402	0.368	0.423	0.423
lCl	0.394	0.391	0.392	0.349	0.412	0.412
uCl	0.406	0.413	0.412	0.387	0.434	0.434
Sen 1	0.901	0.906	0.908	0.854	0.874	0.872
lCl	0.893	0.885	0.888	0.831	0.851	0.848
uCl	0.909	0.927	0.928	0.877	0.897	0.896
Sen 2	0.701	0.697	0.695	0.760	0.660	0.654
lCl	0.691	0.680	0.678	0.737	0.642	0.636
uCl	0.711	0.714	0.716	0.783	0.678	0.672
Sen 3	0.651	0.646	0.647	0.719	0.632	0.632
lCl	0.641	0.629	0.629	0.696	0.612	0.613
uCl	0.661	0.663	0.665	0.742	0.652	0.651
Spe 1	0.991	0.996	0.997	0.922	0.998	0.997
lCl	0.988	0.987	0.991	0.897	0.994	0.993
uCl	0.994	1.000	1.000	1.000	1.000	1.000
Spe 2	0.800	0.799	0.798	0.813	0.784	0.783
lCl	0.792	0.785	0.784	0.805	0.770	0.769
uCl	0.808	0.813	0.812	0.831	0.798	0.797
Spe 3	0.851	0.849	0.849	0.868	0.854	0.854
lCl	0.844	0.835	0.835	0.851	0.839	0.839
uCl	0.858	0.863	0.863	0.885	0.869	0.869
η_{12}^+	-0.005	-0.003	0.000 ¹	0.002	0.031	0.033
η_{13}^+	-0.006	-0.002	-0.003	-0.012	0.008	0.008
η_{23}^+	0.118	0.115	0.115	0.095	0.116	0.117
η_{123}^+	0.008	0.009	0.009	0.010	0.008	0.008
η_{12}^-	0.003	0.001	0.000 ¹	0.005	0.000 ¹	0.000 ¹
η_{13}^-	0.003	0.001	0.000 ¹	0.009	0.000 ¹	0.000 ¹
η_{23}^-	0.088	0.092	0.092	0.073	0.1011	0.101
η_{123}^-	0.000 ¹	0.000 ¹	0.000 ¹	0.003	0.000 ¹	0.000 ¹
Iterations	10000	10000	10000	10000	10000	10000

¹by rounding

Table S17: Results of the Bayesian LCA for the well-chosen starting values in the third simulation scenario

Para-meter	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.029	0.028	0.028	0.018	0.022	0.011
lCl	0.026	0.022	0.023	0.011	0.012	0.008
uCl	0.032	0.034	0.033	0.025	0.032	0.014
Sen 1	0.900	0.901	0.901	0.851	0.930	0.792
lCl	0.891	0.876	0.876	0.826	0.922	0.765
uCl	0.909	0.926	0.926	0.876	0.938	0.819

Sen 2	0.700	0.697	0.698	0.752	0.659	0.800
lCl	0.689	0.672	0.673	0.727	0.647	0.775
uCl	0.711	0.752	0.723	0.777	0.671	0.825
Sen 3	0.649	0.647	0.646	0.722	0.619	0.800
lCl	0.637	0.622	0.621	0.697	0.607	0.776
uCl	0.661	0.673	0.672	0.747	0.631	0.826
Spe 1	0.989	0.988	0.988	0.979	0.983	0.967
lCl	0.897	0.983	0.983	0.973	0.974	0.964
uCl	0.991	0.993	0.993	0.985	0.982	0.970
Spe 2	0.800	0.799	0.799	0.797	0.788	0.789
lCl	0.796	0.794	0.794	0.792	0.783	0.785
uCl	0.804	0.804	0.804	0.802	0.793	0.793
Spe 3	0.850	0.849	0.849	0.847	0.840	0.840
lCl	0.846	0.845	0.845	0.842	0.835	0.836
uCl	0.854	0.853	0.853	0.852	0.845	0.844
η_{12}^+	0.016	0.014	0.015	0.039	0.011	0.053
η_{13}^+	0.008	0.009	0.006	0.031	0.006	0.053
η_{23}^+	0.129	0.128	0.127	0.085	0.126	0.056
η_{123}^+	0.010	0.009	0.009	0.000 ¹	0.006	-0.008
η_{12}^-	0.000 ¹	0.001	0.000 ¹	0.004	0.004	0.009
η_{13}^-	0.000 ¹	0.001	0.000 ¹	0.004	0.004	0.009
η_{23}^-	0.086	0.086	0.086	0.087	0.091	0.091
η_{123}^-	0.000 ¹	0.000 ¹	0.000 ¹	-0.002	-0.002	-0.005
Iterations	10000	10000	10000	10000	10000	10000

¹by rounding

Table S18: Results of the Bayesian LCA for the well-chosen starting values in the fourth simulation scenario

Para-meter	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.400	0.400	0.397	0.393	0.365	0.315
lCl	0.393	0.379	0.371	0.383	0.325	0.295
uCl	0.407	0.421	0.423	0.403	0.405	0.335
Sen 1	0.800	0.807	0.815	0.853	0.613	0.787
lCl	0.790	0.763	0.771	0.839	0.574	0.738
uCl	0.810	0.851	0.859	0.876	0.652	0.836
Sen 2	0.660	0.673	0.670	0.716	0.625	0.778
lCl	0.650	0.636	0.637	0.699	0.582	0.728
uCl	0.670	0.710	0.703	0.733	0.668	0.838
Sen 3	0.701	0.717	0.719	0.753	0.615	0.781
lCl	0.691	0.678	0.683	0.742	0.576	0.732
uCl	0.711	0.756	0.755	0.764	0.654	0.830
Spe 1	0.950	0.955	0.956	0.978	0.802	0.849
lCl	0.944	0.926	0.924	0.967	0.773	0.824
uCl	0.956	0.984	0.988	0.989	0.831	0.874
Spe2	0.850	0.859	0.855	0.882	0.803	0.840
lCl	0.842	0.832	0.826	0.870	0.769	0.812
uCl	0.858	0.886	0.029	0.894	0.837	0.868
Spe 3	0.881	0.892	0.890	0.911	0.801	0.844
lCl	0.873	0.866	0.858	0.899	0.771	0.817
uCl	0.889	0.918	0.922	0.923	0.831	0.871
η_{12}^+	0.036	0.016	0.026	-0.016	0.079	0.037
η_{13}^+	0.044	0.032	0.030	-0.009	0.138	0.054

η_{23}^+	0.070	0.074	0.075	0.069	0.112	0.054
η_{123}^+	0.005	0.007	0.005	0.005	-0.009	-0.006
η_{12}^-	0.017	0.021	0.015	0.006	0.070	0.030
η_{13}^-	0.018	0.016	0.015	0.009	0.066	0.045
η_{23}^-	0.057	0.044	0.045	0.019	0.080	0.049
η_{123}^-	-0.004	-0.004	-0.001	-0.002	-0.016	-0.004
Iterations	10000	10000	10000	10000	10000	10000

Table S19: Results of the Bayesian LCA for the well-chosen starting values in the fifth simulation scenario

Para-meter	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.200	0.201	⁻²	0.157	0.180	0.221
lCl	0.196	0.180		0.161	0.166	0.212
uCl	0.204	0.222		0.163	0.194	0.230
Sen 1	0.717	0.700	⁻²	0.772	0.685	0.679
lCl	0.706	0.665		0.761	0.642	0.659
uCl	0.728	0.735		0.783	0.728	0.699
Sen 2	0.651	0.637	⁻²	0.709	0.639	0.623
lCl	0.641	0.602		0.694	0.592	0.595
uCl	0.661	0.672		0.722	0.686	0.651
Sen 3	0.970	0.959	⁻²	0.976	0.955	0.940
lCl	0.965	0.928		0.946	0.937	0.917
uCl	0.975	0.990		0.996	0.973	0.963
Spe 1	0.979	0.976	⁻²	0.958	0.955	0.986
lCl	0.976	0.961		0.950	0.944	0.978
uCl	0.982	0.991		0.966	0.966	0.994
Spe 2	0.991	0.987	⁻²	0.972	0.972	0.998
lCl	0.988	0.974		0.964	0.965	0.995
uCl	0.994	1.000		0.980	0.979	1.000
Spe 3	0.980	0.978	⁻²	0.936	0.954	0.996
lCl	0.976	0.956		0.925	0.940	0.988
uCl	0.984	1.000		0.947	0.968	1.000
η_{12}^+	0.127	0.131	⁻²	0.124	0.155	0.126
η_{13}^+	0.010	0.009	⁻²	0.009	0.012	0.013
η_{23}^+	0.003	0.008	⁻²	0.008	0.008	0.006
η_{123}^+	0.004	0.007	⁻²	0.000 ¹	0.006	0.013
η_{12}^-	0.002	0.006	⁻²	0.013	0.016	0.000 ¹
η_{13}^-	0.003	0.009	⁻²	0.022	0.027	0.001
η_{23}^-	0.003	0.007	⁻²	0.018	0.022	0.000 ¹
η_{123}^-	0.000 ¹	-0.004	⁻²	-0.010	-0.013	0.000 ¹
Iterations	10000	10000	8000 ³	10000	10000	10000

¹by rounding

²model did not converge and the algorithm aborted the calculation after 8000 Iterations

Table S20: Results of the Bayesian LCA for the poorly chosen starting values in the first simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.791 (0.501, 1.000)	0.262 (0.250, 0.274)	0.263 (0.242, 0.284)
Sen 1 (lCl,uCl)	0.335 (0.186, 0.484)	0.851 (0.836, 0.866)	0.716 (0.690, 0.742)

Sen 2 (lCl,uCl)	0.329 (0.164, 0.658)	0.853 (0.838, 0.868)	0.700 (0.676, 0.724)
Sen 3 (lCl,uCl)	0.317 (0.163, 0.471)	0.850 (0.835, 0.865)	0.821 (0.747, 0.895)
Spe 1 (lCl,uCl)	0.522 (0.257, 0.787)	0.900 (0.889, 0.911)	0.826 (0.804, 0.848)
Spe 2 (lCl,uCl)	0.545 (0.277, 0.813)	0.933 (0.919, 0.947)	0.843 (0.813, 0.873)
Spe 3 (lCl,uCl)	0.555 (0.285, 0.825)	0.932 (0.921, 0.943)	0.897 (0.855, 0.939)
η_{12}^+	0.132	0.040	0.085
η_{13}^+	0.139	0.044	0.092
η_{23}^+	0.133	0.037	0.077
η_{123}^+	0.048	-0.007	-0.028
η_{12}^-	0.036	0.022	0.090
η_{13}^-	0.038	0.036	0.071
η_{23}^-	0.035	0.021	0.069
η_{123}^-	-0.008	-0.006	-0.041
Iterations	10000	10000	10000

Table S21: Results of the Bayesian LCA for the poorly chosen starting values in the second simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.290 (0.155, 0.425)	0.333 (0.315, 0.351)	0.975 (0.970, 0.980)
Sen 1 (lCl,uCl)	0.595 (0.384, 0.806)	0.882 (0.855, 0.909)	0.390 (0.383, 0.397)
Sen 2 (lCl,uCl)	0.764 (0.533, 0.995)	0.813 (0.790, 0.836)	0.427 (0.420, 0.434)
Sen 3 (lCl,uCl)	0.752 (0.537, 0.967)	0.787 (0.763, 0.811)	0.378 (0.370, 0.386)
Spe 1 (lCl,uCl)	0.740 (0.656, 0.824)	0.893 (0.876, 0.910)	0.856 (0.832, 0.880)
Spe 2 (lCl,uCl)	0.766 (0.651, 0.881)	0.810 (0.793, 0.829)	0.928 (0.904, 0.952)
Spe 3 (lCl,uCl)	0.819 (0.701, 0.937)	0.870 (0.855, 0.885)	0.984 (0.956, 1.000)
η_{12}^+	0.012	0.013	0.109
η_{13}^+	0.014	-0.005	0.110
η_{23}^+	0.027	0.048	0.164
η_{123}^+	0.000 ¹	0.004	0.014
η_{12}^-	0.067	-0.006	0.025
η_{13}^-	0.062	-0.005	0.006
η_{23}^-	0.089	0.079	0.005
η_{123}^-	-0.017	0.005	-0.001
Iterations	10000	10000	10000

¹by rounding

Table S22: Results of the Bayesian LCA for the poorly chosen starting values in the third scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence	0.894	0.036	0.027

(lCl,uCl)	(0.815, 0.973)	(0.033,0.039)	(0.021, 0.033)
Sen 1 (lCl,uCl)	0.018 (0.003, 0.033)	0.796 (0.725, 0.867)	0.603 (0.575, 0.631)
Sen 2 (lCl,uCl)	0.181 (0.149, 0.213)	0.759 (0.705, 0.813)	0.898 (0.873, 0.923)
Sen 3 (lCl,uCl)	0.140 (0.111, 0.169)	0.725 (0.670, 0.780)	0.845 (0.810, 0.880)
Spe 1 (lCl,uCl)	0.689 (0.451, 0.927)	0.990 (0.987, 0.993)	0.979 (0.976, 0.982)
Spe 2 (lCl,uCl)	0.413 (0.160, 0.666)	0.804 (0.799, 0.809)	0.806 (0.800, 0.812)
Spe 3 (lCl,uCl)	0.522 (0.262, 0.782)	0.854 (0.850, 0.858)	0.855 (0.849, 0.861)
η_{12}^+	0.005	-0.003	0.008
η_{13}^+	0.005	-0.019	0.021
η_{23}^+	0.082	0.094	0.033
η_{123}^+	0.002	0.012	0.000 ¹
η_{12}^-	0.010	0.000 ¹	0.002
η_{13}^-	0.031	0.000 ¹	0.002
η_{23}^-	0.073	0.083	0.082
η_{123}^-	-0.021	0.000 ¹	0.000 ¹
Iterations	10000	10000	10000

¹by rounding

Table S23: Results of the Bayesian LCA for the poorly chosen starting values in the forth simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.418 (0.130, 0.706)	0.403 (0.385, 0.421)	0.373 (0.364, 0.382)
Sen 1 (lCl,uCl)	0.444 (0.246, 0.642)	0.635 (0.611, 0.659)	0.775 (0.753, 0.797)
Sen 2 (lCl,uCl)	0.402 (0.239, 0.565)	0.639 (0.617, 661)	0.782 (0.760, 0.804)
Sen 3 (lCl,uCl)	0.349 (0.159, 0.539)	0.764 (0.737, 0.791)	0.858 (0.837, 0.879)
Spe 1 (lCl,uCl)	0.634 (0.518, 0.750)	0.847 (0.828, 0.866)	0.901 (0.889, 0.913)
Spe 2 (lCl,uCl)	0.612 (0.489, 0.735)	0.843 (0.824, 0.862)	0.899 (0.885, 0.913)
Spe 3 (lCl,uCl)	0.629 (0.527, 731)	0.927 (0.910, 0.944)	0.948 (0.933, 0.963)
η_{12}^+	0.064	0.083	0.017
η_{13}^+	0.081	0.077	0.004
η_{23}^+	0.079	0.076	0.000 ¹
η_{123}^+	0.016	0.015	0.000 ¹
η_{12}^-	0.088	0.044	0.004
η_{13}^-	0.107	0.036	0.007
η_{23}^-	0.097	0.037	0.009
η_{123}^-	-0.022	-0.011	0.000 ¹
Iterations	10000	10000	10000

¹by rounding

Table S24: Results of the Bayesian LCA for the poorly chosen starting values in the fifth simulation scenario

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.035 (0.007, 0.063)	0.193 (0.183, 0.203)	0.158 (0.154, 0.162)
Sen 1 (lCl,uCl)	0.406 (0.167, 0.645)	0.801 (0.776, 0.826)	0.906 (0.885, 0.927)
Sen 2 (lCl,uCl)	0.351 (0.074, 0.628)	0.724 (0.691, 0.757)	0.825 (0.806, 0.844)
Sen 3 (lCl,uCl)	0.556 (0.313, 0.799)	0.771 (0.752, 0.790)	0.970 (0.745, 0.795)
Spe 1 (lCl,uCl)	0.847 (0.838, 0.856)	0.990 (0.980, 1.000)	0.979 (0.975, 0.983)
Spe 2 (lCl,uCl)	0.865 (0.855, 0.875)	0.998 (0.995, 1.000)	0.990 (0.985, 0.995)
Spe 3 (lCl,uCl)	0.802 (0.787, 0.817)	0.922 (0.912, 0.932)	0.931 (0.926, 0.936)
η_{12}^+	-0.004	0.055	0.009
η_{13}^+	0.023	0.102	0.012
η_{23}^+	0.010	0.126	0.009
η_{123}^+	0.002	-0.037	0.000 ¹
η_{12}^-	0.101	0.001	0.002
η_{13}^-	0.110	0.008	0.004
η_{23}^-	0.103	0.000 ¹	0.003
η_{123}^-	-0.066	0.000 ¹	0.000 ¹
Iterations	10000	10000	10000

¹by rounding

Table S25: Results of the Bayesian LCA for the well-chosen starting values in the third simulation scenario with only positive pairwise dependencies allowed

Para-meter	Starting values 1	Starting values 2	Starting values 3	Starting values 4	Starting values 5	Starting values 6
Prev	0.028	0.024	0.023	0.017	0.019	0.011
lCl	0.026	0.019	0.018	0.012	0.011	0.008
uCl	0.030	0.029	0.028	0.022	0.027	0.014
Sen 1	0.901	0.902	0.903	0.851	0.930	0.791
lCl	0.892	0.877	0.878	0.825	0.922	0.765
uCl	0.910	0.927	0.928	0.877	0.938	0.817
Sen 2	0.698	0.697	0.696	0.751	0.659	0.800
lCl	0.686	0.672	0.671	0.726	0.647	0.775
uCl	0.710	0.722	0.721	0.776	0.671	0.825
Sen 3	0.648	0.648	0.648	0.721	0.619	0.801
lCl	0.636	0.623	0.623	0.696	0.607	0.776
uCl	0.660	0.673	0.673	0.746	0.631	0.826
Spe 1	0.988	0.984	0.984	0.978	0.980	0.967
lCl	0.986	0.980	0.979	0.973	0.973	0.964
uCl	0.990	0.988	0.989	0.983	0.987	0.970
Spe 2	0.799	0.797	0.797	0.796	0.787	0.789
lCl	0.795	0.793	0.792	0.791	0.782	0.785
uCl	0.803	0.801	0.802	0.801	0.792	0.793
Spe 3	0.849	0.847	0.847	0.846	0.839	0.840
lCl	0.845	0.843	0.843	0.841	0.834	0.836
uCl	0.853	0.850	0.850	0.851	0.844	0.844

η_{12}^+	0.027	0.031	0.031	0.054	0.022	0.073
η_{13}^+	0.026	0.029	0.029	0.053	0.021	0.075
η_{23}^+	0.139	0.132	0.136	0.100	0.136	0.070
η_{123}^+	0.004	0.001	0.000 ¹	-0.010	0.002	-0.019
η_{12}^-	0.001	0.003	0.003	0.005	0.005	0.009
η_{13}^-	0.001	0.003	0.003	0.004	0.005	0.009
η_{23}^-	0.086	0.087	0.087	0.088	0.092	0.091
η_{123}^-	0.000 ¹	0.000 ¹	0.000 ¹	-0.002	-0.003	-0.004
Iterations	10000	10000	10000	10000	10000	10000

¹by rounding

Table S26: Results of the Bayesian LCA for the poorly chosen starting values in the third simulation scenario with only positive pairwise dependencies allowed

Parameter	Starting values 1	Starting values 2	Starting values 3
Prevalence (lCl,uCl)	0.711 (0.372, 1.000)	0.034 (0.031, 0.037)	0.024 (0.019, 0.029)
Sen 1 (lCl,uCl)	0.058 (0.000, 0.181)	0.811 (0.716, 0.906)	0.600 (0.573, 0.627)
Sen 2 (lCl,uCl)	0.219 (0.057, 0.381)	0.721 (0.660, 0.782)	0.897 (0.872, 0.922)
Sen 3 (lCl,uCl)	0.185 (0.021, 0.349)	0.686 (0.622, 0.75)	0.846 (0.811, 0.881)
Spe 1 (lCl,uCl)	0.754 (0.483, 1.000)	0.988 (0.984, 0.992)	0.977 (0.974, 0.980)
Spe 2 (lCl,uCl)	0.552 (0.334, 0.770)	0.801 (0.796, 0.806)	0.804 (0.799, 0.809)
Spe 3 (lCl,uCl)	0.615 (0.386, 0.844)	0.851 (0.847, 0.855)	0.853 (0.848, 0.858)
η_{12}^+	0.009	0.026	0.029
η_{13}^+	0.010	0.021	0.044
η_{23}^+	0.073	0.119	0.044
η_{123}^+	0.003	0.007	-0.011
η_{12}^-	0.028	0.000 ¹	0.003
η_{13}^-	0.030	0.000 ¹	0.003
η_{23}^-	0.101	0.085	0.083
η_{123}^-	-0.007	0.000 ¹	0.000 ¹
Iterations	10000	10000	10000

¹by rounding

Anhang 2:

Der folgende Anhang entspricht den Supporting Information der Publikation II (Schoneberg C, Böttcher J, Janowetz B, Rostalski A, Kreienbrock L, Campe A (2022) An intercomparison study of ELISAs for the detection of porcine reproductive and respiratory syndrome virus – evaluating six conditionally dependent tests. PLoS ONE 17(1): e0262944. <https://doi.org/10.1371/journal.pone.0262944>).

S1 Table. Starting values for the stepwise latent class algorithm for the complete data set

Parameter	Starting values MI	Starting values MR	Starting values PI	Starting values PR	Starting values RI	Starting values RR
Prevalence	0.70	0.70	0.72	0.72	0.74	0.74
Sensitivity 1	0.988	0.988	0.92	0.92	0.92	0.92
Sensitivity 2	0.999	0.999	0.92	0.92	0.90	0.90
Sensitivity 3	0.991	0.991	0.90	0.90	0.90	0.90
Sensitivity 4	0.999	0.999	0.96	0.96	0.90	0.90
Sensitivity 5	0.999	0.999	0.85	0.85	0.90	0.90
Sensitivity 6	0.999	0.999	0.95	0.95	0.99	0.99
Specificity 1	0.999	0.999	0.90	0.90	0.97	0.97
Specificity 2	0.998	0.998	0.98	0.98	0.97	0.97
Specificity 3	0.999	0.999	0.95	0.95	0.97	0.97
Specificity 4	0.999	0.999	0.99	0.99	0.97	0.97
Specificity 5	0.972	0.972	0.90	0.90	0.97	0.97
Specificity 6	0.999	0.999	0.95	0.95	0.70	0.70
η_{12}^{+1}	0.000	0.007	0.000	0.005	0.000	0.003
η_{13}^{+1}	0.000	0.012	0.000	0.006	0.000	0.010
η_{14}^{+1}	0.000	0.008	0.000	0.005	0.000	0.007
η_{15}^{+1}	0.000	0.015	0.000	0.015	0.000	0.018
η_{16}^{+1}	0.000	0.030	0.000	0.030	0.000	0.031
η_{23}^{+1}	0.000	0.009	0.000	0.010	0.000	0.008
η_{24}^{+1}	0.000	0.008	0.000	0.010	0.000	0.008
η_{25}^{+1}	0.000	0.010	0.000	0.010	0.000	0.010
η_{26}^{+1}	0.000	0.013	0.000	0.015	0.000	0.014
η_{34}^{+1}	0.000	0.007	0.000	0.005	0.000	0.005
η_{35}^{+1}	0.000	0.008	0.000	0.008	0.000	0.008
η_{36}^{+1}	0.000	0.011	0.000	0.010	0.000	0.011
η_{45}^{+1}	0.000	0.001	0.000	0.001	0.000	0.001
η_{46}^{+1}	0.000	0.008	0.000	0.008	0.000	0.008
η_{56}^{+1}	0.000	0.032	0.000	0.028	0.000	0.028
η_{123}^{+1}	0.000	0.008	0.000	0.005	0.000	0.005
η_{124}^{+1}	0.000	0.007	0.000	0.007	0.000	0.007
η_{125}^{+1}	0.000	0.001	0.000	0.011	0.000	0.011
η_{126}^{+1}	0.000	0.009	0.000	0.008	0.000	0.008
η_{134}^{+1}	0.000	0.006	0.000	0.006	0.000	0.006
η_{135}^{+1}	0.000	0.007	0.000	0.007	0.000	0.007
η_{136}^{+1}	0.000	0.007	0.000	0.007	0.000	0.007
η_{145}^{+1}	0.000	0.006	0.000	0.006	0.000	0.006
η_{146}^{+1}	0.000	0.005	0.000	0.005	0.000	0.005
η_{156}^{+1}	0.000	0.010	0.000	0.010	0.000	0.010
η_{234}^{+1}	0.000	0.006	0.000	0.006	0.000	0.006
η_{235}^{+1}	0.000	0.007	0.000	0.007	0.000	0.007
η_{236}^{+1}	0.000	0.006	0.000	0.006	0.000	0.006
η_{245}^{+1}	0.000	0.007	0.000	0.007	0.000	0.007
η_{246}^{+1}	0.000	0.005	0.000	0.005	0.000	0.005
η_{256}^{+1}	0.000	0.007	0.000	0.007	0.000	0.007
η_{345}^{+1}	0.000	0.006	0.000	0.006	0.000	0.006
η_{346}^{+1}	0.000	0.005	0.000	0.005	0.000	0.005

$\eta_{356}^+{}^1$	0.000	0.006	0.000	0.006	0.000	0.006
$\eta_{456}^+{}^1$	0.000	0.007	0.000	0.007	0.000	0.007
$\eta_{1234}^+{}^1$	0.000	0.006	0.000	0.006	0.000	0.006
$\eta_{1235}^+{}^1$	0.000	0.006	0.000	0.006	0.000	0.006
$\eta_{1236}^+{}^1$	0.000	0.006	0.000	0.006	0.000	0.006
$\eta_{1245}^+{}^1$	0.000	0.006	0.000	0.006	0.000	0.006
$\eta_{1246}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{1256}^+{}^1$	0.000	0.006	0.000	0.006	0.000	0.006
$\eta_{1345}^+{}^1$	0.000	0.006	0.000	0.006	0.000	0.006
$\eta_{1346}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{1356}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{1456}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{2345}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{2346}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{2356}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{2456}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{3456}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{12345}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{12346}^+{}^1$	0.000	0.004	0.000	0.004	0.000	0.004
$\eta_{12356}^+{}^1$	0.000	0.005	0.000	0.005	0.000	0.005
$\eta_{12456}^+{}^1$	0.000	0.004	0.000	0.004	0.000	0.004
$\eta_{13456}^+{}^1$	0.000	0.004	0.000	0.004	0.000	0.004
$\eta_{23456}^+{}^1$	0.000	0.004	0.000	0.004	0.000	0.004
$\eta_{123456}^+{}^1$	0.000	0.004	0.000	0.004	0.000	0.004
$\eta_{12}^-{}^1$	0.000	0.070	0.000	0.050	0.000	0.070
$\eta_{13}^-{}^1$	0.000	0.085	0.000	0.050	0.000	0.050
$\eta_{14}^-{}^1$	0.000	0.072	0.000	0.060	0.000	0.065
$\eta_{15}^-{}^1$	0.000	0.050	0.000	0.050	0.000	0.050
$\eta_{16}^-{}^1$	0.000	0.029	0.000	0.030	0.000	0.029
$\eta_{23}^-{}^1$	0.000	0.102	0.000	0.100	0.000	0.100
$\eta_{24}^-{}^1$	0.000	0.099	0.000	0.100	0.000	0.099
$\eta_{25}^-{}^1$	0.000	0.058	0.000	0.060	0.000	0.058
$\eta_{26}^-{}^1$	0.000	0.028	0.000	0.030	0.000	0.028
$\eta_{34}^-{}^1$	0.000	0.122	0.000	0.120	0.000	0.122
$\eta_{35}^-{}^1$	0.000	0.056	0.000	0.055	0.000	0.056
$\eta_{36}^-{}^1$	0.000	0.034	0.000	0.035	0.000	0.034
$\eta_{45}^-{}^1$	0.000	0.067	0.000	0.065	0.000	0.067
$\eta_{46}^-{}^1$	0.000	0.030	0.000	0.030	0.000	0.030
$\eta_{56}^-{}^1$	0.000	0.021	0.000	0.020	0.000	0.020
$\eta_{123}^-{}^1$	0.000	0.049	0.000	0.050	0.000	0.055
$\eta_{124}^-{}^1$	0.000	0.045	0.000	0.050	0.000	0.045
$\eta_{125}^-{}^1$	0.000	0.033	0.000	0.035	0.000	0.033
$\eta_{126}^-{}^1$	0.000	0.020	0.000	0.020	0.000	0.020
$\eta_{134}^-{}^1$	0.000	0.041	0.000	0.040	0.000	0.041
$\eta_{135}^-{}^1$	0.000	0.031	0.000	0.030	0.000	0.031
$\eta_{136}^-{}^1$	0.000	0.020	0.000	0.020	0.000	0.020
$\eta_{145}^-{}^1$	0.000	0.031	0.000	0.020	0.000	0.025

η_{146}^-	0.000	0.019	0.000	0.020	0.000	0.019
η_{156}^-	0.000	0.017	0.000	0.017	0.000	0.017
η_{234}^-	0.000	0.053	0.000	0.053	0.000	0.053
η_{235}^-	0.000	0.033	0.000	0.030	0.000	0.033
η_{236}^-	0.000	0.017	0.000	0.017	0.000	0.017
η_{245}^-	0.000	0.035	0.000	0.035	0.000	0.035
η_{246}^-	0.000	0.017	0.000	0.017	0.000	0.017
η_{256}^-	0.000	0.015	0.000	0.015	0.000	0.015
η_{345}^-	0.000	0.032	0.000	0.032	0.000	0.032
η_{346}^-	0.000	0.018	0.000	0.018	0.000	0.018
η_{356}^-	0.000	0.013	0.000	0.013	0.000	0.013
η_{456}^-	0.000	0.013	0.000	0.013	0.000	0.013
η_{1234}^-	0.000	0.037	0.000	0.036	0.000	0.036
η_{1235}^-	0.000	0.026	0.000	0.026	0.000	0.026
η_{1236}^-	0.000	0.017	0.000	0.017	0.000	0.017
η_{1245}^-	0.000	0.026	0.000	0.026	0.000	0.026
η_{1246}^-	0.000	0.016	0.000	0.016	0.000	0.016
η_{1256}^-	0.000	0.014	0.000	0.012	0.000	0.012
η_{1345}^-	0.000	0.026	0.000	0.026	0.000	0.026
η_{1346}^-	0.000	0.016	0.000	0.015	0.000	0.016
η_{1356}^-	0.000	0.014	0.000	0.014	0.000	0.014
η_{1456}^-	0.000	0.013	0.000	0.013	0.000	0.013
η_{2345}^-	0.000	0.027	0.000	0.027	0.000	0.027
η_{2346}^-	0.000	0.014	0.000	0.013	0.000	0.013
η_{2356}^-	0.000	0.012	0.000	0.012	0.000	0.012
η_{2456}^-	0.000	0.012	0.000	0.012	0.000	0.012
η_{3456}^-	0.000	0.011	0.000	0.015	0.000	0.015
η_{12345}^-	0.000	0.020	0.000	0.020	0.000	0.020
η_{12346}^-	0.000	0.013	0.000	0.013	0.000	0.013
η_{12356}^-	0.000	0.011	0.000	0.011	0.000	0.011
η_{12456}^-	0.000	0.010	0.000	0.010	0.000	0.010
η_{13456}^-	0.000	0.010	0.000	0.010	0.000	0.010
η_{23456}^-	0.000	0.009	0.000	0.009	0.000	0.009
η_{123456}^-	0.000	0.008	0.000	0.008	0.000	0.008

¹ η_{ij}^+ is the dependency of the sensitivities of test i and test j; η_{ij}^- is the dependency of the specificities of test i and test j

S2 Table. Starting values for the stepwise latent class algorithm for the vaccinated subgroup

Parameter	Starting values MI	Starting values MR	Starting values RI	Starting values RR
Prevalence	0.87	0.87	0.84	0.84
Sensitivity 1	0.988	0.988	0.92	0.92
Sensitivity 2	0.999	0.999	0.90	0.90
Sensitivity 3	0.991	0.991	0.90	0.90
Sensitivity 4	0.999	0.999	0.90	0.90
Sensitivity 5	0.999	0.999	0.90	0.90
Sensitivity 6	0.999	0.999	0.99	0.99

Specificity 1	0.999	0.999	0.97	0.97
Specificity 2	0.998	0.998	0.97	0.97
Specificity 3	0.999	0.999	0.97	0.97
Specificity 4	0.999	0.999	0.97	0.97
Specificity 5	0.972	0.972	0.97	0.97
Specificity 6	0.999	0.999	0.70	0.70
η_{12}^{+1}	0.000	0.016	0.000	0.016
η_{13}^{+1}	0.000	0.012	0.000	0.012
η_{14}^{+1}	0.000	0.008	0.000	0.008
η_{15}^{+1}	0.000	0.021	0.000	0.021
η_{16}^{+1}	0.000	0.030	0.000	0.030
η_{23}^{+1}	0.000	0.009	0.000	0.009
η_{24}^{+1}	0.000	0.008	0.000	0.008
η_{25}^{+1}	0.000	0.010	0.000	0.010
η_{26}^{+1}	0.000	0.013	0.000	0.013
η_{34}^{+1}	0.000	0.007	0.000	0.007
η_{35}^{+1}	0.000	0.008	0.000	0.008
η_{36}^{+1}	0.000	0.011	0.000	0.011
η_{45}^{+1}	0.000	0.001	0.000	0.001
η_{46}^{+1}	0.000	0.008	0.000	0.008
η_{56}^{+1}	0.000	0.032	0.000	0.032
η_{123}^{+1}	0.000	0.008	0.000	0.008
η_{124}^{+1}	0.000	0.007	0.000	0.007
η_{125}^{+1}	0.000	0.011	0.000	0.011
η_{126}^{+1}	0.000	0.009	0.000	0.009
η_{134}^{+1}	0.000	0.006	0.000	0.006
η_{135}^{+1}	0.000	0.007	0.000	0.007
η_{136}^{+1}	0.000	0.007	0.000	0.007
η_{145}^{+1}	0.000	0.006	0.000	0.006
η_{146}^{+1}	0.000	0.005	0.000	0.005
η_{156}^{+1}	0.000	0.010	0.000	0.010
η_{234}^{+1}	0.000	0.006	0.000	0.006
η_{235}^{+1}	0.000	0.007	0.000	0.007
η_{236}^{+1}	0.000	0.006	0.000	0.006
η_{245}^{+1}	0.000	0.007	0.000	0.007
η_{246}^{+1}	0.000	0.005	0.000	0.005
η_{256}^{+1}	0.000	0.007	0.000	0.007
η_{345}^{+1}	0.000	0.006	0.000	0.006
η_{346}^{+1}	0.000	0.005	0.000	0.005
η_{356}^{+1}	0.000	0.006	0.000	0.006
η_{456}^{+1}	0.000	0.007	0.000	0.007
η_{1234}^{+1}	0.000	0.006	0.000	0.006
η_{1235}^{+1}	0.000	0.006	0.000	0.006
η_{1236}^{+1}	0.000	0.006	0.000	0.006
η_{1245}^{+1}	0.000	0.006	0.000	0.006
η_{1246}^{+1}	0.000	0.005	0.000	0.005
η_{1256}^{+1}	0.000	0.006	0.000	0.006
η_{1345}^{+1}	0.000	0.006	0.000	0.006

η_{1346}^+	0.000	0.005	0.000	0.005
η_{1356}^+	0.000	0.005	0.000	0.005
η_{1456}^+	0.000	0.005	0.000	0.005
η_{2345}^+	0.000	0.005	0.000	0.005
η_{2346}^+	0.000	0.005	0.000	0.005
η_{2356}^+	0.000	0.005	0.000	0.005
η_{2456}^+	0.000	0.005	0.000	0.005
η_{3456}^+	0.000	0.005	0.000	0.005
η_{12345}^+	0.000	0.005	0.000	0.005
η_{12346}^+	0.000	0.004	0.000	0.004
η_{12356}^+	0.000	0.005	0.000	0.005
η_{12456}^+	0.000	0.004	0.000	0.004
η_{13456}^+	0.000	0.004	0.000	0.004
η_{23456}^+	0.000	0.004	0.000	0.004
η_{123456}^+	0.000	0.004	0.000	0.004
η_{12}^-	0.000	0.081	0.000	0.081
η_{13}^-	0.000	0.085	0.000	0.085
η_{14}^-	0.000	0.072	0.000	0.072
η_{15}^-	0.000	0.050	0.000	0.050
η_{16}^-	0.000	0.029	0.000	0.029
η_{23}^-	0.000	0.102	0.000	0.102
η_{24}^-	0.000	0.099	0.000	0.099
η_{25}^-	0.000	0.058	0.000	0.058
η_{26}^-	0.000	0.028	0.000	0.028
η_{34}^-	0.000	0.122	0.000	0.122
η_{35}^-	0.000	0.056	0.000	0.056
η_{36}^-	0.000	0.034	0.000	0.034
η_{45}^-	0.000	0.067	0.000	0.067
η_{46}^-	0.000	0.030	0.000	0.030
η_{56}^-	0.000	0.021	0.000	0.021
η_{123}^-	0.000	0.049	0.000	0.049
η_{124}^-	0.000	0.045	0.000	0.045
η_{125}^-	0.000	0.033	0.000	0.033
η_{126}^-	0.000	0.020	0.000	0.020
η_{134}^-	0.000	0.041	0.000	0.041
η_{135}^-	0.000	0.031	0.000	0.031
η_{136}^-	0.000	0.020	0.000	0.020
η_{145}^-	0.000	0.031	0.000	0.031
η_{146}^-	0.000	0.019	0.000	0.019
η_{156}^-	0.000	0.017	0.000	0.017
η_{234}^-	0.000	0.053	0.000	0.053
η_{235}^-	0.000	0.033	0.000	0.033
η_{236}^-	0.000	0.017	0.000	0.017
η_{245}^-	0.000	0.035	0.000	0.035
η_{246}^-	0.000	0.017	0.000	0.017
η_{256}^-	0.000	0.015	0.000	0.015
η_{345}^-	0.000	0.032	0.000	0.032

η_{346}^-	0.000	0.018	0.000	0.018
η_{356}^-	0.000	0.013	0.000	0.013
η_{456}^-	0.000	0.013	0.000	0.013
η_{1234}^-	0.000	0.037	0.000	0.037
η_{1235}^-	0.000	0.026	0.000	0.026
η_{1236}^-	0.000	0.017	0.000	0.017
η_{1245}^-	0.000	0.026	0.000	0.026
η_{1246}^-	0.000	0.016	0.000	0.016
η_{1256}^-	0.000	0.014	0.000	0.014
η_{1345}^-	0.000	0.026	0.000	0.026
η_{1346}^-	0.000	0.016	0.000	0.016
η_{1356}^-	0.000	0.014	0.000	0.014
η_{1456}^-	0.000	0.013	0.000	0.013
η_{2345}^-	0.000	0.027	0.000	0.027
η_{2346}^-	0.000	0.014	0.000	0.014
η_{2356}^-	0.000	0.012	0.000	0.012
η_{2456}^-	0.000	0.012	0.000	0.012
η_{3456}^-	0.000	0.011	0.000	0.011
η_{12345}^-	0.000	0.020	0.000	0.020
η_{12346}^-	0.000	0.013	0.000	0.013
η_{12356}^-	0.000	0.011	0.000	0.011
η_{12456}^-	0.000	0.010	0.000	0.010
η_{13456}^-	0.000	0.010	0.000	0.010
η_{23456}^-	0.000	0.009	0.000	0.009
η_{123456}^-	0.000	0.008	0.000	0.008

¹ η_{ij}^+ is the dependency of the sensitivities of test i and test j; η_{ij}^- is the dependency of the specificities of test i and test j

S3 Table. Resulting values for the stepwise latent class algorithm for the complete data set for the six starting value sets

Parameter	Starting values MI	Starting values MR	Starting values PI	Starting values PR	Starting values RI	Starting values RR
Prevalence	0.7545134	0.7666076	0.7567326	0.7679239	0.7529343	0.7637762
	0.7247642	0.7373689	0.7270746	0.7387427	0.7231208	0.7344151
	0.7842626	0.7958463	0.7863905	0.7971050	0.7827478	0.7931373
Sensitivity 1	0.8885030	0.8767316	0.8864680	0.8756724	0.8899512	0.8792851
	0.8667465	0.8540074	0.8645391	0.8528646	0.8683189	0.8567648
	0.9102595	0.8994557	0.9083970	0.8984801	0.9115836	0.9018053
Sensitivity 2	0.8436503	0.8309218	0.8415248	0.8297584	0.8451777	0.8335559
	0.8185454	0.8050127	0.8162818	0.8037785	0.8201732	0.8078087
	0.8687551	0.8568309	0.8667679	0.8557382	0.8701822	0.8593031
Sensitivity 3	0.7969744	0.7845615	0.7948701	0.7833689	0.7984812	0.7871817
	0.7691692	0.7561428	0.7669581	0.7548934	0.7707532	0.7588893
	0.8247796	0.8129801	0.8227820	0.8118444	0.8262092	0.8154741
Sensitivity 4	0.7561575	0.7450837	0.7540996	0.7438616	0.7576277	0.7476597
	0.7264758	0.7149585	0.7243335	0.7136891	0.7280069	0.7176354
	0.7858392	0.7752088	0.7838657	0.7740341	0.7872486	0.7776841
Sensitivity 5	0.9129749	0.9048626	0.9112788	0.9036336	0.9141841	0.9070473

lCl	0.8934908	0.8845814	0.8916240	0.8832356	0.8948231	0.8869761
uCl	0.9324589	0.9251439	0.9309335	0.9240316	0.9335452	0.9271186
Sensitivity 6	0.9525908	0.9498199	0.9513969	0.9490386	0.9534354	0.9513975
lCl	0.9379011	0.9347291	0.9365327	0.9338370	0.9388707	0.9365334
uCl	0.9672805	0.9649108	0.9662611	0.9642403	0.9680001	0.9662616
Specificity 1	0.9492908	0.9540399	0.9506034	0.9552470	0.9483496	0.9523384
lCl	0.9341248	0.9395654	0.9356246	0.9409549	0.9330511	0.9376116
uCl	0.9644568	0.9685143	0.9655822	0.9695392	0.9636481	0.9670652
Specificity 2	0.9684986	0.9687751	0.9692956	0.9694612	0.9679626	0.9677067
lCl	0.9564249	0.9567527	0.9573707	0.9575674	0.9557899	0.9554871
uCl	0.9805724	0.9807974	0.9812206	0.9813550	0.9801353	0.9799263
Specificity 3	0.9821027	0.9817019	0.9826638	0.9821019	0.9817152	0.9809894
lCl	0.9729384	0.9724375	0.9736417	0.9729374	0.9724540	0.9715496
uCl	0.9912670	0.9909664	0.9916859	0.9912664	0.9909764	0.9904291
Specificity 4	0.9883819	0.9905897	0.9887722	0.9907187	0.9881036	0.9901010
lCl	0.9809746	0.9839158	0.9814889	0.9840903	0.9806091	0.9832577
uCl	0.9957891	0.9972635	0.9960554	0.9973471	0.9955980	0.9969443
Specificity 5	0.8775750	0.8918949	0.8795105	0.8923472	0.8762073	0.8894086
lCl	0.8549178	0.8704310	0.8570084	0.8709229	0.8534418	0.8677296
uCl	0.9002321	0.9133587	0.9020126	0.9137716	0.8989729	0.9110876
Specificity 6	0.7713395	0.7997516	0.7742294	0.8014178	0.7692863	0.7958679
lCl	0.7423096	0.7720892	0.7453295	0.7738420	0.7401652	0.7680064
uCl	0.8003695	0.8274140	0.8031293	0.8289935	0.7984075	0.8237294
η_{12}^{+1}	0.0160583	0.0157224	0.0155491	0.0152712	0.0164139	0.0164217
η_{13}^{+1}	0.0127707	0.0130732	0.0123663	0.0127620	0.0130512	0.0136039
η_{14}^{+1}	0.0078364	0.0071215	0.0075776	0.0069914	0.0080154	0.0074637
η_{15}^{+1}	0.0215780	0.0195980	0.0210145	0.0191953	0.0219715	0.0203212
η_{16}^{+1}	0.0308721	0.0289343	0.0301886	0.0281790	0.0313640	0.0298910
η_{23}^{+1}	0.0089938	0.0090419	0.0086850	0.0088119	0.0092127	0.0094733
η_{24}^{+1}	0.0080675	0.0072983	0.0078000	0.0071637	0.0082591	0.0076593
η_{25}^{+1}	0.0141282	0.0136593	0.0137060	0.0133439	0.0144215	0.0142361
η_{26}^{+1}	0.0128971	0.0133050	0.0125348	0.0129251	0.0131537	0.0138397
η_{34}^{+1}	0.0071134	0.0066447	0.0068761	0.0065212	0.0072823	0.0069677
η_{35}^{+1}	0.0080491	0.0079772	0.0077849	0.0078062	0.0082302	0.0083386
η_{36}^{+1}	0.0116525	0.0122630	0.0113183	0.0120174	0.0118839	0.0127084
η_{45}^{+1}	0.0101958	0.0083930	0.0098750	0.0082822	0.0104238	0.0088043
η_{46}^{+1}	0.0083727	0.0070406	0.0081279	0.0069618	0.0085494	0.0073655
η_{56}^{+1}	0.0329482	0.0295674	0.0322496	0.0291915	0.0334719	0.0306336
η_{123}^{+1}	0.0083832	0.0084755	0.0081127	0.0082766	0.0085741	0.0088549
η_{124}^{+1}	0.0068689	0.0063724	0.0066559	0.0062621	0.0070204	0.0066644
η_{125}^{+1}	0.0108430	0.0105979	0.0105335	0.0103551	0.0110550	0.0110236
η_{126}^{+1}	0.0094003	0.0097685	0.0091566	0.0095190	0.0095705	0.0101277
η_{134}^{+1}	0.0064745	0.0061039	0.0062711	0.0059971	0.0066185	0.0063835
η_{135}^{+1}	0.0070052	0.0069975	0.0067916	0.0068550	0.0071505	0.0072941
η_{136}^{+1}	0.0072547	0.0076639	0.0070510	0.0075010	0.0073970	0.0079525
η_{145}^{+1}	0.0066263	0.0061672	0.0064286	0.0060606	0.0067618	0.0064371
η_{146}^{+1}	0.0051269	0.0049601	0.0049907	0.0048864	0.0052217	0.0051559
η_{156}^{+1}	0.0097529	0.0095597	0.0095638	0.0093995	0.0098804	0.0098247
η_{234}^{+1}	0.0063089	0.0059637	0.0061123	0.0058559	0.0064540	0.0062402
η_{235}^{+1}	0.0069300	0.0069143	0.0067142	0.0067643	0.0070818	0.0072185

η_{236}^{+1}	0.0066345	0.0069127	0.0064395	0.0067571	0.0067749	0.0071968
η_{245}^{+1}	0.0069156	0.0063808	0.0067063	0.0062687	0.0070651	0.0066714
η_{246}^{+1}	0.0055026	0.0052415	0.0053506	0.0051593	0.0056140	0.0054626
η_{256}^{+1}	0.0068492	0.0070009	0.0066869	0.0068607	0.0069614	0.0072409
η_{345}^{+1}	0.0061491	0.0058513	0.0059602	0.0057467	0.0062829	0.0061146
η_{346}^{+1}	0.0052223	0.0050585	0.0050743	0.0049785	0.0053301	0.0052715
η_{356}^{+1}	0.0057426	0.0059460	0.0055897	0.0058393	0.0058478	0.0061641
η_{456}^{+1}	0.0069649	0.0060012	0.0067865	0.0059414	0.0070928	0.0062477
η_{1234}^{+1}	0.0060074	0.0057037	0.0058277	0.0056075	0.0061400	0.0059583
η_{1235}^{+1}	0.0066324	0.0066490	0.0064333	0.0065123	0.0067721	0.0069311
η_{1236}^{+1}	0.0064515	0.0067580	0.0062670	0.0066114	0.0065837	0.0070275
η_{1245}^{+1}	0.0060657	0.0057158	0.0058892	0.0056183	0.0061909	0.0059628
η_{1246}^{+1}	0.0047916	0.0046742	0.0046649	0.0046026	0.0048841	0.0048605
η_{1256}^{+1}	0.0064590	0.0066298	0.0062985	0.0064930	0.0065698	0.0068653
η_{1345}^{+1}	0.0057306	0.0054879	0.0055614	0.0053936	0.0058502	0.0057251
η_{1346}^{+1}	0.0049015	0.0047741	0.0047674	0.0047014	0.0049990	0.0049678
η_{1356}^{+1}	0.0056442	0.0058104	0.0054900	0.0057019	0.0057510	0.0060306
η_{1456}^{+1}	0.0047007	0.0045768	0.0045801	0.0045065	0.0047847	0.0047516
η_{2345}^{+1}	0.0055477	0.0053298	0.0053862	0.0052357	0.0056668	0.0055620
η_{2346}^{+1}	0.0047447	0.0046366	0.0046171	0.0045638	0.0048422	0.0048265
η_{2356}^{+1}	0.0052913	0.0054652	0.0051466	0.0053573	0.0053952	0.0056788
η_{2456}^{+1}	0.0048809	0.0047206	0.0047546	0.0046473	0.0049733	0.0056788
η_{3456}^{+1}	0.0046350	0.0045586	0.0045120	0.0044879	0.0047246	0.0047394
η_{12345}^{+1}	0.0052804	0.0050957	0.0051334	0.0050119	0.0053887	0.0053087
η_{12346}^{+1}	0.0045151	0.0044321	0.0043995	0.0043680	0.0046036	0.0046058
η_{12356}^{+1}	0.0050298	0.0052213	0.0048992	0.0051249	0.0051234	0.0054154
η_{12456}^{+1}	0.0042258	0.0041885	0.0041225	0.0041257	0.0043011	0.0043442
η_{13456}^{+1}	0.0043098	0.0042676	0.0042010	0.0042045	0.0043889	0.0044288
η_{23456}^{+1}	0.0041700	0.0041420	0.0040666	0.0040788	0.0042492	0.0042999
η_{123456}^{-1}	0.0039697	0.0039606	0.0038762	0.0039050	0.0040414	0.0041047
η_{12}^{-1}	0.0804363	0.0888623	0.0818446	0.0895466	0.0794234	0.0871423
η_{13}^{-1}	0.0836392	0.0915566	0.0849911	0.0922259	0.0826717	0.0899027
η_{14}^{-1}	0.0705120	0.0778084	0.0718170	0.0784721	0.0695807	0.0762048
η_{15}^{-1}	0.0485143	0.0542293	0.0497000	0.0550206	0.0476705	0.0527314
η_{16}^{-1}	0.0283202	0.0301515	0.0292386	0.0307519	0.0276676	0.0289379
η_{23}^{-1}	0.1009448	0.1093372	0.1022793	0.1100281	0.0999770	0.1077262
η_{24}^{-1}	0.0978313	0.1054391	0.0991301	0.1061349	0.0968913	0.1038674
η_{25}^{-1}	0.0563866	0.0622800	0.0575692	0.0631081	0.0555339	0.0607915
η_{26}^{-1}	0.0272217	0.0290233	0.0280770	0.0295793	0.0266073	0.0278969
η_{34}^{-1}	0.1202998	0.1272379	0.1215088	0.1279145	0.1194288	0.1257664
η_{35}^{-1}	0.0553636	0.0609566	0.0564846	0.0617613	0.0545598	0.0595365
η_{36}^{-1}	0.0335389	0.0352238	0.0343786	0.0357656	0.0329413	0.0341226
η_{45}^{-1}	0.0658047	0.0708853	0.0669047	0.0716833	0.0650165	0.0694970
η_{46}^{-1}	0.0294466	0.0310463	0.0302588	0.0315735	0.0288687	0.0299829
η_{56}^{-1}	0.0204275	0.0205425	0.0211785	0.0211904	0.0198847	0.0194923
η_{123}^{-1}	0.0484458	0.0519566	0.0490241	0.0521798	0.0480207	0.0513233
η_{124}^{-1}	0.0446394	0.0478864	0.0452097	0.0481181	0.0442220	0.0472536

η_{125}^- ¹	0.0325730	0.0354490	0.0331894	0.0358442	0.0321240	0.0347171
η_{126}^- ¹	0.0194242	0.0200989	0.0199577	0.0204462	0.0190360	0.0194060
η_{134}^- ¹	0.0404228	0.0430980	0.0409152	0.0432997	0.0400649	0.0425563
η_{135}^- ¹	0.0307727	0.0333907	0.0313551	0.0337725	0.0303525	0.0326988
η_{136}^- ¹	0.0190785	0.0195724	0.0195931	0.0199068	0.0187091	0.0189092
η_{145}^- ¹	0.0305481	0.0330029	0.0311068	0.0333656	0.0301454	0.0323417
η_{146}^- ¹	0.0187855	0.0193427	0.0192841	0.0196637	0.0184272	0.0186991
η_{156}^- ¹	0.0160575	0.0156599	0.0165772	0.0161285	0.0156763	0.0149306
η_{234}^- ¹	0.0526338	0.0550674	0.0530182	0.0552208	0.0523454	0.0546738
η_{235}^- ¹	0.0322814	0.0348486	0.0327967	0.0352051	0.0319015	0.0342427
η_{236}^- ¹	0.0165794	0.0170389	0.0170110	0.0173115	0.0162641	0.0164843
η_{245}^- ¹	0.0341234	0.0363726	0.0346180	0.0367159	0.0337592	0.0357966
η_{246}^- ¹	0.0169468	0.0174061	0.0173680	0.0176725	0.0166391	0.0168669
η_{256}^- ¹	0.0146511	0.0141594	0.0151085	0.0145775	0.0143103	0.0135131
η_{345}^- ¹	0.0313945	0.0333127	0.0318146	0.0336152	0.0310871	0.0328227
η_{346}^- ¹	0.0178064	0.0180837	0.0181825	0.0183192	0.0175352	0.0176049
η_{356}^- ¹	0.0130313	0.0125064	0.0134565	0.0128926	0.0127185	0.0119065
η_{456}^- ¹	0.0128839	0.0123522	0.0132861	0.0127201	0.0125878	0.0117866
η_{1234}^- ¹	0.0364099	0.0390985	0.0368649	0.0392825	0.0360770	0.0385927
η_{1235}^- ¹	0.0259208	0.0281754	0.0263915	0.0284794	0.0255765	0.0276185
η_{1236}^- ¹	0.0162598	0.0166903	0.0166811	0.0169641	0.0159527	0.0161445
η_{1245}^- ¹	0.0260387	0.0281557	0.0264929	0.0284452	0.0257069	0.0276204
η_{1246}^- ¹	0.0153434	0.0157627	0.0157375	0.0160175	0.0150557	0.0152513
η_{1256}^- ¹	0.0130928	0.0126735	0.0134954	0.0130438	0.0127918	0.0121013
η_{1345}^- ¹	0.0252551	0.0272686	0.0257011	0.0275605	0.0249327	0.0267398
η_{1346}^- ¹	0.0157147	0.0160625	0.0161101	0.0163171	0.0154302	0.0155528
η_{1356}^- ¹	0.0132412	0.0128136	0.0136446	0.0131796	0.0129444	0.0122469
η_{1456}^- ¹	0.0123708	0.0119651	0.0127462	0.0123069	0.0120943	0.0114360
η_{2345}^- ¹	0.0264720	0.0284765	0.0268711	0.0287535	0.0261775	0.0280065
η_{2346}^- ¹	0.0139494	0.0143204	0.0142909	0.0145350	0.0136999	0.0138806
η_{2356}^- ¹	0.0114671	0.0110494	0.0118194	0.0113723	0.0112036	0.0105473
η_{2456}^- ¹	0.0113951	0.0109513	0.0117309	0.0112609	0.0111439	0.0104742
η_{3456}^- ¹	0.0105111	0.0100947	0.0108344	0.0103888	0.0102728	0.0096370
η_{12345}^- ¹	0.0196196	0.0209563	0.0199111	0.0211396	0.0194037	0.0206288
η_{12346}^- ¹	0.0122733	0.0123793	0.0125513	0.0125577	0.0120687	0.0120239
η_{12356}^- ¹	0.0105267	0.0100189	0.0108197	0.0102937	0.0103058	0.0096025
η_{12456}^- ¹	0.0098432	0.0093679	0.0101159	0.0096249	0.0096371	0.0089786
η_{13456}^- ¹	0.0099003	0.0094065	0.0101710	0.0096584	0.0096992	0.0090250
η_{23456}^- ¹	0.0085480	0.0080781	0.0087832	0.0082997	0.0083699	0.0077412
η_{123456}^- ¹	0.0079618	0.0074703	0.0081616	0.0076630	0.0078094	0.0071838
Log-Likelihood	-1402.758	-1402.758	-1402.758	-1402.758	-1402.758	-1402.758
Iterations ²	13	11	12	10	11	10

lCl: lower confidence limit, uCl: upper confidence limit

¹ η_{ij}^+ is the dependency of the sensitivities of test i and test j; η_{ij}^- is the dependency of the specificities of test i and test j

² The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

S4 Table. Resulting values for the stepwise latent class algorithm for the vaccinated subgroup for the four starting value sets

Parameter	Starting values MI	Starting values MR	Starting values RI	Starting values RR
Prevalence	0.8340011	0.8797012	0.6874968	0.8738472
lCl	0.8002903	0.8502278	0.6455020	0.8437658
uCl	0.8677119	0.9091747	0.7294916	0.9039287
Sensitivity 1	0.9177018	0.8791367	0.9995038	0.8836254
lCl	0.8928029	0.8496036	0.9974862	0.8545721
uCl	0.9426006	0.9086698	1.0015215	0.9126788
Sensitivity 2	0.8761292	0.8337009	0.9823997	0.8392046
lCl	0.8462822	0.7999657	0.9704863	0.8059230
uCl	0.9059763	0.8674361	0.9943132	0.8724862
Sensitivity 3	0.8415881	0.8011879	0.9738739	0.8064909
lCl	0.8085072	0.7650284	0.9594221	0.7706991
uCl	0.8746690	0.8373473	0.9883257	0.8422827
Sensitivity 4	0.8224948	0.7807156	0.9545927	0.7857367
lCl	0.7878765	0.7432284	0.9357300	0.7485622
uCl	0.8571130	0.8182029	0.9734555	0.8229112
Sensitivity 5	0.9455100	0.9132076	1.0027371	0.9177181
lCl	0.9249452	0.8877006	- ³	0.8928215
uCl	0.9660748	0.9387145	- ³	0.9426147
Sensitivity 6	0.9898129	0.9687037	1.0021484	0.9609851
lCl	0.9807151	0.9529285	- ³	0.9434419
uCl	0.9989107	0.9844789	- ³	0.9785282
Specificity 1	0.9252195	0.9634234	0.7100122	0.9554180
lCl	0.9013881	0.9464158	0.6689014	0.9367193
uCl	0.9490509	0.9804310	0.7511230	0.9741166
Specificity 2	0.9609231	0.9686463	0.8022970	0.9695379
lCl	0.9433666	0.9528571	0.7662136	0.9539676
uCl	0.9784796	0.9844355	0.8383804	0.9851081
Specificity 3	0.9675930	0.9795592	0.8792660	0.9800632
lCl	0.9515495	0.9667389	0.8497465	0.9673987
uCl	0.9836365	0.9923795	0.9087854	0.9927277
Specificity 4	0.9875143	0.9897116	0.8983858	0.9887411
lCl	0.9774540	0.9805692	0.8710116	0.9791819
uCl	0.9975746	0.9988540	0.9257600	0.9983003
Specificity 5	0.8332342	0.8928543	0.5940496	0.8866938
lCl	0.7994611	0.8648316	0.5495577	0.8579762
uCl	0.8670072	0.9208771	0.6385415	0.9154113
Specificity 6	0.7211439	0.8368640	0.4149786	0.7460164
lCl	0.6805151	0.8033879	0.3703378	0.7065788
uCl	0.7617727	0.8703402	0.4596194	0.7854539
η_{12}^{+1}	0.0159569	0.0108192	0.0980023	0.0143303
η_{13}^{+1}	0.0228745	0.0132577	0.0738307	0.0154019
η_{14}^{+1}	0.0059857	0.0066310	0.0380955	0.0079797
η_{15}^{+1}	0.0220352	0.0140534	0.1130476	0.0190629
η_{16}^{+1}	0.0464414	0.0254506	0.1150169	0.0290739
η_{23}^{+1}	0.0072372	0.0066216	0.0429163	0.0078226
η_{24}^{+1}	0.0064075	0.0069675	0.0496910	0.0081000
η_{25}^{+1}	0.0130288	0.0091876	0.0905796	0.0127718

η_{26}^{+1}	0.0175709	0.0129860	0.0753081	0.0139250
η_{34}^{+1}	0.0058696	0.0082833	0.0497412	0.0078978
η_{35}^{+1}	0.0067213	0.0079095	0.0461127	0.0079197
η_{36}^{+1}	0.0158840	0.0119472	0.0447802	0.0106882
η_{45}^{+1}	0.0104035	0.0091860	0.0603639	0.0099832
η_{46}^{+1}	0.0090040	0.0086100	0.0421677	0.0083993
η_{56}^{+1}	0.0657057	0.0353451	0.1282003	0.0351269
η_{123}^{+1}	0.0064567	0.0063600	0.0209900	0.0072992
η_{124}^{+1}	0.0046992	0.0055926	0.0165465	0.0069074
η_{125}^{+1}	0.0102060	0.0084210	0.0303162	0.0106337
η_{126}^{+1}	0.0104858	0.0087410	0.0198826	0.0101974
η_{134}^{+1}	0.0048919	0.0055037	0.0268194	0.0071698
η_{135}^{+1}	0.0062566	0.0056152	0.0188130	0.0073419
η_{136}^{+1}	0.0080645	0.0056043	0.0129731	0.0069120
η_{145}^{+1}	0.0047123	0.0057759	0.0111434	0.0068609
η_{146}^{+1}	0.0037367	0.0052874	0.0041215	0.0056257
η_{156}^{+1}	0.0122255	0.0110208	0.0118962	0.0131294
η_{234}^{+1}	0.0045518	0.0054894	0.0258303	0.0068677
η_{235}^{+1}	0.0064006	0.0060486	0.0196214	0.0069789
η_{236}^{+1}	0.0049422	0.0054377	0.0097692	0.0056857
η_{245}^{+1}	0.0051763	0.0061265	0.0203148	0.0070384
η_{246}^{+1}	0.0044014	0.0056973	0.0129683	0.0058860
η_{256}^{+1}	0.0085973	0.0080006	0.0149769	0.0094550
η_{345}^{+1}	0.0048068	0.0073144	0.0248630	0.0069138
η_{346}^{+1}	0.0040345	0.0068091	0.0160912	0.0057715
η_{356}^{+1}	0.0039661	0.0064495	0.0058181	0.0056821
η_{456}^{+1}	0.0066821	0.0073238	0.0120228	0.0070521
η_{1234}^{+1}	0.0042647	0.0053849	0.0213724	0.0065847
η_{1235}^{+1}	0.0059525	0.0058900	0.0180911	0.0066943
η_{1236}^{+1}	0.0049600	0.0054847	0.0119670	0.0055770
η_{1245}^{+1}	0.0039541	0.0050061	0.0126647	0.0061482
η_{1246}^{+1}	0.0034016	0.0046835	0.0089786	0.0051586
η_{1256}^{+1}	0.0073709	0.0070185	0.0187106	0.0079217
η_{1345}^{+1}	0.0042360	0.0049781	0.0184771	0.0064285
η_{1346}^{+1}	0.0036330	0.0046624	0.0129245	0.0054003
η_{1356}^{+1}	0.0057857	0.0052469	0.0135686	0.0059520
η_{1456}^{+1}	0.0034440	0.0048413	0.0081824	0.0051307
η_{2345}^{+1}	0.0037881	0.0048998	0.0159412	0.0060891
η_{2346}^{+1}	0.0032860	0.0045950	0.0115360	0.0051251
η_{2356}^{+1}	0.0045928	0.0050381	0.0104453	0.0051934
η_{2456}^{+1}	0.0037428	0.0051238	0.0107279	0.0052516
η_{3456}^{+1}	0.0034774	0.0061229	0.0116709	0.0051603
η_{12345}^{+1}	0.0035464	0.0048058	0.0121250	0.0058361
η_{12346}^{+1}	0.0030717	0.0045048	0.0085382	0.0049108
η_{12356}^{+1}	0.0042471	0.0049130	0.0064641	0.0049805
η_{12456}^{+1}	0.0028514	0.0041898	0.0046327	0.0045868
η_{13456}^{+1}	0.0030388	0.0041603	0.0070821	0.0047904

η_{23456}^{+1}	0.0027321	0.0041007	0.0063813	0.0045427
η_{123456}^{+1}	0.0025580	0.0040219	0.0051376	0.0043540
η_{12}^{-1}	0.0581897	0.0864124	-0.0020096	0.0826783
η_{13}^{-1}	0.0655521	0.0928564	-0.0003273	0.0894910
η_{14}^{-1}	0.0534255	0.0802348	-0.0026961	0.0771694
η_{15}^{-1}	0.0349634	0.0569859	-0.0037895	0.0536542
η_{16}^{-1}	0.0082767	0.0257903	-0.0028363	0.0326335
η_{23}^{-1}	0.0731372	0.1010337	-0.0018495	0.0971081
η_{24}^{-1}	0.0773753	0.1059528	0.0001910	0.1022971
η_{25}^{-1}	0.0413796	0.0652149	-0.0029091	0.0613260
η_{26}^{-1}	0.0084748	0.0254751	-0.0021574	0.0315668
η_{34}^{-1}	0.1059088	0.1311717	0.0118620	0.1280478
η_{35}^{-1}	0.0396468	0.0612354	-0.0034569	0.0579875
η_{36}^{-1}	0.0075013	0.0233508	-0.0027812	0.0296242
η_{45}^{-1}	0.0448178	0.0677602	-0.0026128	0.0646519
η_{46}^{-1}	0.0083788	0.0244335	-0.0020508	0.0306555
η_{56}^{-1}	0.0051186	0.0214348	-0.0025019	0.0282652
η_{123}^{-1}	0.0397030	0.0537100	-0.0027525	0.0517160
η_{124}^{-1}	0.0364333	0.0500165	-0.0027578	0.0483095
η_{125}^{-1}	0.0260142	0.0383834	-0.0037974	0.0359810
η_{126}^{-1}	0.0062453	0.0182549	-0.0028375	0.0231225
η_{134}^{-1}	0.0356712	0.0474235	-0.0026184	0.0463139
η_{135}^{-1}	0.0251306	0.0367225	-0.0037146	0.0349264
η_{136}^{-1}	0.0063860	0.0179651	-0.0027625	0.0230414
η_{145}^{-1}	0.0258460	0.0375261	-0.0036248	0.0358084
η_{146}^{-1}	0.0062633	0.0176238	-0.0027133	0.0226306
η_{156}^{-1}	0.0033275	0.0154868	-0.0031996	0.0211930
η_{234}^{-1}	0.0441101	0.0545899	-0.0025578	0.0533265
η_{235}^{-1}	0.0253974	0.0365593	-0.0036155	0.0344427
η_{236}^{-1}	0.0054480	0.0158115	-0.0027819	0.0201246
η_{245}^{-1}	0.0298184	0.0417184	-0.0027765	0.0397689
η_{246}^{-1}	0.0061823	0.0165729	-0.0020591	0.0208120
η_{256}^{-1}	0.0036371	0.0152661	-0.0025111	0.0202495
η_{345}^{-1}	0.0268383	0.0364227	-0.0032674	0.0350269
η_{346}^{-1}	0.0050909	0.0141251	-0.0026294	0.0182810
η_{356}^{-1}	0.0028902	0.0136829	-0.0031348	0.0188437
η_{456}^{-1}	0.0037534	0.0146138	-0.0023939	0.0196866
η_{1234}^{-1}	0.0318870	0.0428960	-0.0026714	0.0414996
η_{1235}^{-1}	0.0217656	0.0313707	-0.0037232	0.0294257
η_{1236}^{-1}	0.0054277	0.0152852	-0.0027647	0.0194117
η_{1245}^{-1}	0.0220192	0.0314808	-0.0036273	0.0296419
η_{1246}^{-1}	0.0052532	0.0147979	-0.0027105	0.0188179
η_{1256}^{-1}	0.0025716	0.0127019	-0.0031904	0.0171855
η_{1345}^{-1}	0.0216640	0.0308458	-0.0035523	0.0294179
η_{1346}^{-1}	0.0054383	0.0148103	-0.0026420	0.0190410
η_{1356}^{-1}	0.0029276	0.0128776	-0.0031157	0.0176142
η_{1456}^{-1}	0.0027665	0.0123070	-0.0030543	0.0168998

η_{2345}^- ¹	0.0217635	0.0306505	-0.0034535	0.0289533
η_{2346}^- ¹	0.0046444	0.0131045	-0.0026572	0.0167090
η_{2356}^- ¹	0.0022690	0.0114099	-0.0031258	0.0154990
η_{2456}^- ¹	0.0030345	0.0122063	-0.0023971	0.0162392
η_{3456}^- ¹	0.0024360	0.0110388	-0.0029924	0.0152172
η_{12345}^- ¹	0.0179440	0.0244076	-0.0035543	0.0230497
η_{12346}^- ¹	0.0044721	0.0119031	-0.0026393	0.0152187
η_{12356}^- ¹	0.0022386	0.0103158	-0.0031068	0.0140091
η_{12456}^- ¹	0.0021088	0.0098692	-0.0030456	0.0134498
η_{13456}^- ¹	0.0023978	0.0099857	-0.0029743	0.0137630
η_{23456}^- ¹	0.0018573	0.0088419	-0.0029839	0.0121044
η_{123456}^- ¹	0.0018408	0.0080563	-0.0029657	0.0110102
Log-Likelihood	-691.0331	-691.0331	-691.0331	-691.0331
Iterations ²	11	11	16	24

ICL: lower confidence limit, uCL: upper confidence limit

¹ η_{ij}^+ is the dependency of the sensitivities of test i and test j; η_{ij}^- is the dependency of the specificities of test i and test j

² The number of the iterations the whole algorithm, not the ones of the EM algorithm performed at each step

³Confidence limits could not be calculated

S5 Table. Studies evaluating the diagnostic test accuracies of the tests analyzed in this publication that were used as sources of information for suitable starting values for the iterative, frequentist latent class analysis

Source	Sample	Tests used in the source and in our study	Sensitivities	Specificities
Díaz et al. (2012)	Blood samples from experimentally and naturally infected pigs	PRRS X3	100%	-
Gerber et al. (2014)	Oral fluid samples from experimentally infected boars	PRRS X3	100%	100%
Sipos et al. (2009)	Serum samples collected during routine diagnostics on pig farms	PRRS X3	100% (assumed gold standard)	100% (assumed gold standard)
		Ingezim PRRS universal	82.2-97.3%	71.1-92.3%
Sattler et al. (2016)	Serum samples of piglets vaccinated with an inactivated PRRSV vaccine and subsequent challenge	PRRS X3	90% (Type 1 vaccine) 100% (field strain)	-
		pigtyp PRRSV Ab	90% (Type 1 vaccine) 100% (field strain)	-
		Ingezim PRRS 2.0	100%	-

Sattler et al. (2015)	Serum samples of PRRSV positive and negative herds, pigs of monitored PRRSV negative herds, herds with unknown herd history and wild boars	PRRS X3	100% (assumed gold standard)	100% (assumed gold standard)
		INgezim PRRS 2.0	96.0%	99.0%
		PrioCHECK	100%	95.1%
Sattler et al. (2014)	Serum samples of pigs after PRRSV type 2 live vaccination and subsequent challenge	PRRS X3	-	100%
		pigtyp PRRSV Ab	Higher sensitivity than PRRS X3	98,1%
		Ingezim PRRS 2.0	Higher than PRRS X3	Higher than Ingezim PRRS universal
Karniychuk and Nauwynck (2014)	Serum from Belarusian swine farms	PRRS X3	80.0%	83.0%
		Ingezim PRRS universal	83.0%	77.0%
Biernacka et al. (2018)	Serum samples obtained from Polish pig farms	PRRS X3	64.8%	100%
		INgezim PRRS 2.0	59.3%	96.6%
		Pigtype PRRSV Ab	63.6%	98.6%
		PrioCHECK	67.2%	94.2%
Henao-Diaz et al. (2019)	Oral fluid samples with known status collected under experimental conditions	PRRS X3	100%	100%
		pigtyp PRRSV Ab	94.0%	100%

Danksagung

An dieser Stelle möchte ich mich bei all jenen bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Zunächst gilt mein ganz besonderer Dank Herrn Prof. Dr. Lothar Kreienbrock für die hervorragende Begleitung und Betreuung meiner Doktorarbeit.

Ferner danke ich Frau Prof. Dr. Katja Ickstadt und Herrn Prof. Dr. Martin Ganter für ihre Unterstützung und fachliche Beratung bei der Erstellung dieser Arbeit.

Bei Frau PD Dr. Amely Campe möchte ich mich für die wissenschaftliche Unterstützung während meiner Promotionszeit, die vielen guten Ratschläge sowie die Hilfe bei der Themenfindung bedanken.

Auch möchte ich mich bei Herrn Prof. Dr. Hans-Peter Piepho für die Bereitschaft der Übernahme des Zweitgutachtens bedanken.

Ein großer Dank geht an all meinen Kolleginnen und Kollegen am Institut für Biometrie, Epidemiologie und Informationsverarbeitung für den wissenschaftlichen Austausch, die Hilfsbereitschaft und die angenehme Arbeitsatmosphäre.

Den Projektpartnern im Forschungsverbund Q-GAPS möchte ich für die Unterstützung bei fachlichen Fragen und den spannenden Austausch danken. Ein besonderer Dank gilt dabei Fenja Winter für die großartige Teamarbeit und ihre freundschaftliche Unterstützung im Laufe des Projektes.

Außerdem danke ich dem Bayrischen Tiergesundheitsdienst für die Bereitstellung der Daten für Publikation 2. Insbesondere Herrn Dr. Jens Böttcher möchte ich für die fachliche Unterstützung und die spannenden Diskussionen im Rahmen der Erstellung der Publikation danken.

Abschließend möchte ich mich bei und meiner Familie, besonders bei meinem Mann Niels, für ihre Unterstützung in vielerlei Hinsicht bedanken. Danke, dass ihr immer an mich glaubt und mich jederzeit ermutigt habt!